# Project 1

Using a dataset with Yelp scores in Madison, Wisconsin, model the probability of 1-star, 2-star, 3-star, 4-star, and 5-star reviews as a function of the two included neighborhoods. Summarize the model parameters and make a graphic / table (preferably a graphic) to display your results.

Note these dataset is filtered to only include 2500 individual reviews. You can remove this step (and would for a standard data analysis), but be prepared to give the algorithm more time to execute.

```
set.seed(02132022)
yelp_biz <- read_csv("https://math.montana.edu/ahoegh/teaching/stat532/data/yelp_biz_info.csv") %>%
  filter(neighborhood %in% c("South Campus",'Williamson - Marquette', "Capitol"))
yelp_reviews <- read_csv("https://math.montana.edu/ahoegh/teaching/stat532/data/yelp_biz_reviews.csv")

yelp_comb <- yelp_reviews %>%
  right_join(yelp_biz, by = 'business_id') %>%
  mutate(stars = factor(stars)) %>%
  sample_n(2500) %>%
  mutate(pizza = str_detect(categories, 'Pizza'), coffee = str_detect(categories, 'Coffee'))
```

**Part 1** Create at least two data visualizations that compare the star ranking with `neighborhood`, `pizza`, and `coffee`. The figures should include informative titles, legends, and captions.

**Part 2** Write out the statistical model you will use for the ordinal response. Include complete notation.

**Part 3** Fit the model and interpret the model outcome.

**Part 4** Include graphics to illustrate the model results. In particular, show the expected probability of 1 - 5 stars of the following cases:

- `neighborhood` = Capitol, pizza = T, coffee = F
- `neighborhood` = Williamson - Marquette, pizza = F, coffee = F
- `neighborhood` = South Campus

**Part 5** Write a paragraph to a friend that attends the University of Wisconsin giving them recommendations about what businesses to visit.