Project 2 **Math 222**
Exploratory Data Analysis **Due Friday, February 27**

1. The way that questions are worded in a survey might affect how people respond. For example, students in one introductory statistics class were randomly assigned to answer one of the following questions:

   - Should your college allow speeches on campus that might incite violence?
   - Should your college forbid speeches on campus that might incite violence?

   Of the 11 students who received the first question, 8 responded yes. Of the 14 students who received the second question, 12 said no.

   (a) Identify the observational units and the explanatory and response variables.

   (b) Is this an observational study or an experiment? If an observational study, suggest a potential confounding variable. If an experiment, describe whether you think confounding variables have been adequately controlled.

   (c) Construct a two-way table in R to summarize these results.

   (d) Construct a segmented bar graph to display these results and comment on the relationship revealed by this graph.

   (e) Based on earlier studies, researchers expected people to be less likely to agree to "forbid" the speeches, leading to more no responses (and thus appearing to be in favor of having the speeches), whereas they expected people to be comparatively less likely to agree to "allow" the speeches. Do these data provide strong evidence that these students responded more positively toward having such speeches if their question was phrased in terms of "forbid" rather than "allow"? Use R to carry out a test of significance and explain the decision you would make based on the p-value. Write a paragraph summarizing your conclusions including whether a cause-and-effect conclusion can be drawn and the population you are willing to generalize these results to.

   (f) In a 1976 study, one group of subjects was asked, "Do you think the United States should forbid public speeches in favor of communism?", whereas another group was asked, "Do you think the United States should allow public speeches in favor of communism?". Of the 409 subjects randomly asked the "forbid" version, 161 favored the forbidding of communist speeches. Of the 432 subjects asked the "allow" version, 189 favored allowing the speeches. Construct a segmented bar graph for these data and comment on whether you believe the p-value for this table will be larger or smaller than that in (e). Explain your reasoning.

2. Myopia, or near-sightedness, typically develops during the childhood years. Recent studies have explored whether there is an association between development of myopia and the use of night-lights with infants. Quinn, Shin, Maguire, and Stone (1999) examined the type of light children aged 2-16 were exposed to. Between January and June 1998, the parents of 479 children who were seen as outpatients in a university pediatric ophthalmology clinic completed a questionnaire (children who had already developed serious

eye conditions were excluded). One of the questions asked was "Under which lighting condition did/does your child sleep at night?" before the age of 2 years. The following two-way table classifies the children's eye condition and whether or not they slept with some kind of light (e.g., a night light or full room light) or in darkness.

```
---- Example R Code -----------
vision <- matrix(c(40, 39, 12, 114, 115, 22, 18, 78, 22),
    ncol = 3,
    byrow = TRUE)
colnames(vision) <- c("Dark", "Night-light", "Room-light")
rownames(vision) <- c("Far-sighted", "Normal", "Near-sighted")
vision <- as.table(vision)
vision
------------------------------
```

	Dark	Night-light	Room-light
Far-sighted	40	39	12
Normal	114	115	22
Near-sighted	18	78	22

(a) Which variable, lighting condition or eye condition, would you consider the explanatory variable in this study and which the response variable?

(b) Use R to make a segmented bar graph to that shows the percent of kids with each eye condition (with separate bars for each lighting level). Write a few words to describe any relationship or trend that you see in the graph.

(c) Carry out a χ^2-test for association to see if the relationship between light at night and near-sightedness is statistically significant.

(d) Make a 95% confidence interval for the difference in the proportions of kids who are near-sighted between the kids who sleep in darkness versus those who sleep with some kind of light at night. Write a sentence or two to explain what your confidence interval means.