Biscuits are big business: Amalgamated Biscuit Company (ABC) sells biscuits in nearly 10,000 advertising markets in the United States. As the head of advertising for ABC, you've been asked by the CEO to analyze the efficacy of different media (Internet, TV, radio and newspapers) and to make recommendations on next year's advertising budget. Unfortunately, the company does a better job making biscuits than they do keeping advertising and sales records, so you have data from only 200 out of the 10,000 markets. From what you can tell, though, these 200 markets seem to be a random and representative sample of the 10,000. All of these markets run independent advertisement campaigns.

For each of the 200 markets, you know the amount of money spent on advertising on the Internet, TV, radio and in newspapers. These data are in the first four columns of *advert.csv*. The *AREA* column indicates whether the market is in the north or the south of the country. The *SALES* column gives the annual sales of ABC biscuits listed in the sixth column. (All numbers are in units of $1000 USD). The *MEDAGE* column gives the median age for that market, and the *MHI* column gives the median household income for that market.

To help you prepare for the final exam, work through the following tasks:

* Compute summary statistics for each of the quantitative variables in the data frame.
* Create histograms for each of the quantitative variables in the data frame.
* Plot the relationship between biscuit sales and age.
  + Is a linear model appropriate for this data?
    - If no, explain which conditions are violated and discuss what, if anything, can be done to fix this.
    - If yes, fit the linear model, and evaluate it by looking at a histogram of the residuals and a plot of residuals vs fitted values.
  + Repeat this for only markets in the south.
* Is there a significant difference in income between the north and the south? Explain your answer.
* What proportion of the markets spend more on internet ads than on any other type of ads?
* What proportion of the markets spend less on television ads than on any other type of ads?
* What is the correlation between spending on radio ads and spending on newspaper adds? Is this correlation significantly different from 0? How do you know?
* Fit a linear model of income as a function of age. Are the conditions of linear regression satisfied? How do you know? Is the slope of your model significantly different from 0? How do you know?