In-class R activity

```
# Load the Galton height data from the `HistData` library.
install.packages("HistData")
library("HistData")
data(GaltonFamilies)
str(GaltonFamilies)
```

Questions:

- How many children are there in the data?
- How many unique families are there?
- What are the overall mean & median height of the children?
- What are the overall mean & median height of the male children? Idem female children?
- Can you visualize the distributions of children's height by gender?
- Use a linear model to estimate the mean height of children.
- Use a linear model to estimate the mean height of each gender. What do the diagnostic plots tell you about this model?
- Fit a multiple regression model predicting children's height from father's height, mother's height, and gender.
- Change reference level for the gender variable to "female" and re-estimate the model. Inspect the residuals.
- Scale the parents' height, re-estimate the model, and compare interpretations.
- We would expect siblings to be somewhat similar in height as they share genetic factors through their parents and environmental factors through their shared upbringing. Fit a new model to estimate both the population means as well as how average family heights vary around these population means. Hint: mixed-effects models can be used for this.
- Fit a fixed-effects-only model to control for family and compare interpretations.

R hints:

- lm: built-in
- lmer: `library(lme4); library(lmerTest)`
  - Recall typing speed example: `lmer(errors ~ speed + (1 | ID), data = data)`
- vif: `library(car)`