Many situations come down to comparing the means of two groups. There are at least three different data collection strategies that result in such a problem.

1. We identify two different populations or processes and choose a random sample from each.
   For example, we might compare Kellogg’s Raisin Bran and Meijer’s Raisin Bran to determine which kind of cereal has more raisins. We would need a random sample of boxes of each to answer this question.

2. We identify one population and choose a random sample from it. We record two different variables – a binary categorical variable that divides the sample into two groups and a quantitative variable.
   For example, we might take a random sample of Calvin students, measure their heights, and compare the average height of the males and females of the students in the sample.

3. We perform a randomized comparative experiment where there are two treatments and one quantitative response variable.
   For example, we heat some rubberbands and leave others at ambient temperature and compare how the stretchability of the rubber bands in each treatment group.

```r
> mean(stretch ~ treat, rubberbands)

A  H
244.09 253.50

> dotplot(stretch ~ treat, rubberbands)
> lm(stretch ~ treat, rubberbands)
```

```
Call:
  lm(formula = stretch ~ treat, data = rubberbands)

Coefficients:
(Intercept)   treatH
  244.09       9.41
```

```
```

```r
> stretch <- c(220, 230, 240, 250, 260, 270)
> treat <- c("A", "H")

plot(stretch ~ treat, pch = 19, cex = 2)
```