

Basic Statistics for Variables in a Dataset

- Most data in R come in the form of a *data frame*, which we will refer to as a *dataset*.
- It is simply a rectangular arrangement of data, where each row corresponds to an individual or case that we study, and each column corresponds to one kind of measurement.
- We most typically call these columns *variables*.
- Most of the standard functions provide a formula¹ interface. Some examples:

```
data(counties)
View(counties)
favstats(~pop2010, data=counties)    # Five number summary
mean(~pop2010, data=counties)        # Average population for counties
median(~pop2010, data=counties)      # Median population for counties
median(~pop2010|state, data=counties) # Median county population by state
```

¹[formulas.html](#)