Basic Statistics for Variables in Datasets

- 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 a specific question about the cases.
- In statistics, we typically call these columns *variables*.
- Most of the basic functions provide a formula[1] interface. You will need the hanoverbase package[2] in order to use this 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
[2]packagesAsToolboxes.html