## Describing and Learning About Relationships

- Statistics is not just about describing a single group of numbers.
- Powerful way to describe and discover relationships from noisy data.

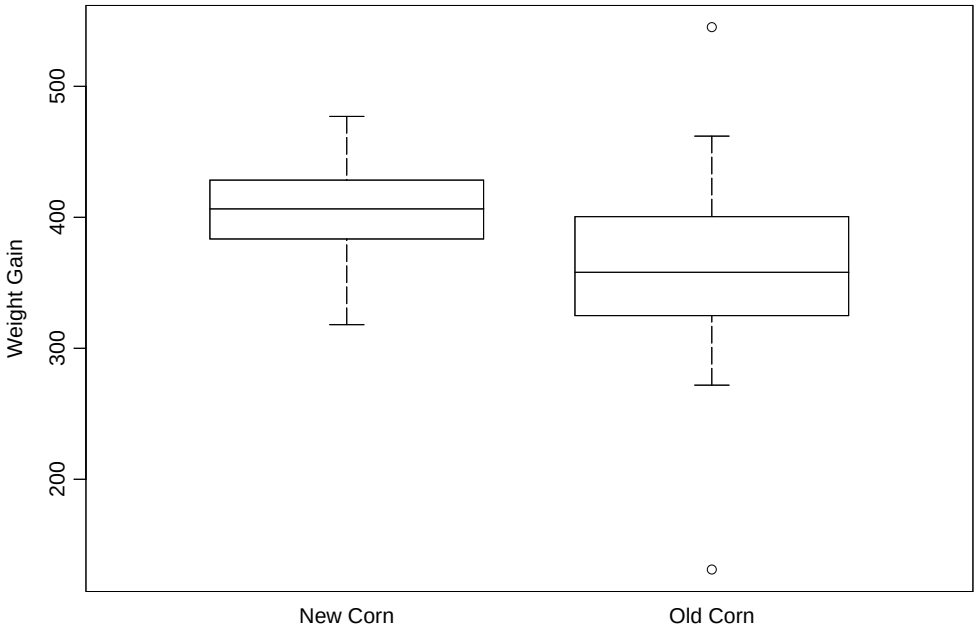
### *Example: Chick weights*

Corn is an important animal food. Normal corn lacks certain amino acids, which are building blocks for protein. Plant scientists have developed new corn varieties that have more of these amino acids. To test a new corn as an animal food, a group of 20 one-day-old male chicks was fed a ration containing the new corn. A control group of another 20 chicks was fed a ration that was identical except that it contained normal corn. Here are the weight gains (in grams) after 21 days.

| Normal corn | | | | New corn | | | |
|---|---|---|---|---|---|---|---|
| 380 | 321 | 366 | 356 | 361 | 447 | 401 | 375 |
| 283 | 349 | 402 | 462 | 434 | 403 | 393 | 426 |
| 356 | 410 | 329 | 399 | 406 | 318 | 467 | 407 |
| 350 | 384 | 316 | 272 | 427 | 420 | 477 | 392 |
| 545 | 455 | 360 | 131 | 430 | 339 | 410 | 326 |

*Some terminology*

- The corn type (normal or new) is the *independent variable*. Other names for this include:
  - predictor
  - stimulus
  - explanatory variable
  - $X$ variable
- The weight gain is called the *dependent variable*. Other names for this include:
  - response
  - outcome
  - $Y$ variable



*In* **Experiments** *and* **Comparative Studies***:*

- The conditions being compared are the *independent variable*, whose values are usually *qualitative*: the names of the conditions.
- The *outcome* of the experiment is usually the *dependent variable*. It is usually (but not always) *quantitative:* a count or numerical measurement.
- Useful tools: Parallel boxplots, back to back stem and leaf, etc.

## Other kinds of studies and data

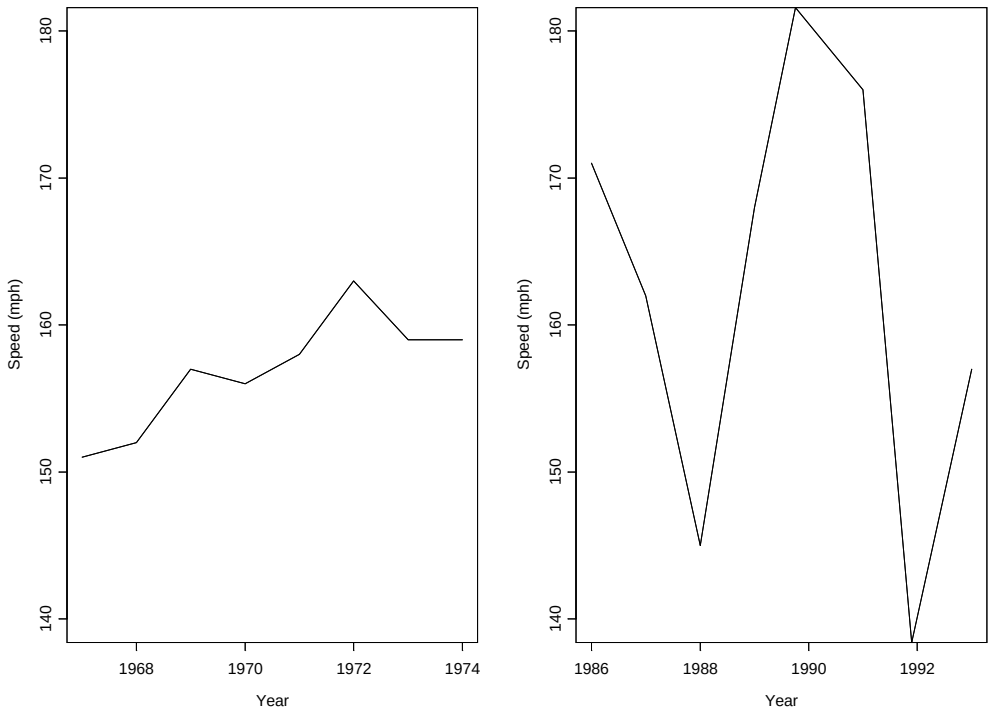
*Time Series*

- Independent variable is *time*
- Response is usually quantitative (counts, amounts, measurements)
- Useful tools: line plots

*Time Series Example*

Average Speeds of Winning Cars, Indy 500, in Two Eras.

| Era 1 | | Era 2 | |
|---|---|---|---|
| YEAR | SPEED | YEAR | SPEED |
| 1967 | 151 | 1986 | 171 |
| 1968 | 152 | 1987 | 162 |
| 1969 | 157 | 1988 | 145 |
| 1970 | 156 | 1989 | 168 |
| 197l | 158 | 1990 | 186 |
| 1972 | 163 | 1991 | 176 |
| 1973 | 159 | 1992 | 134 |
| 1974 | 159 | 1993 | 157 |



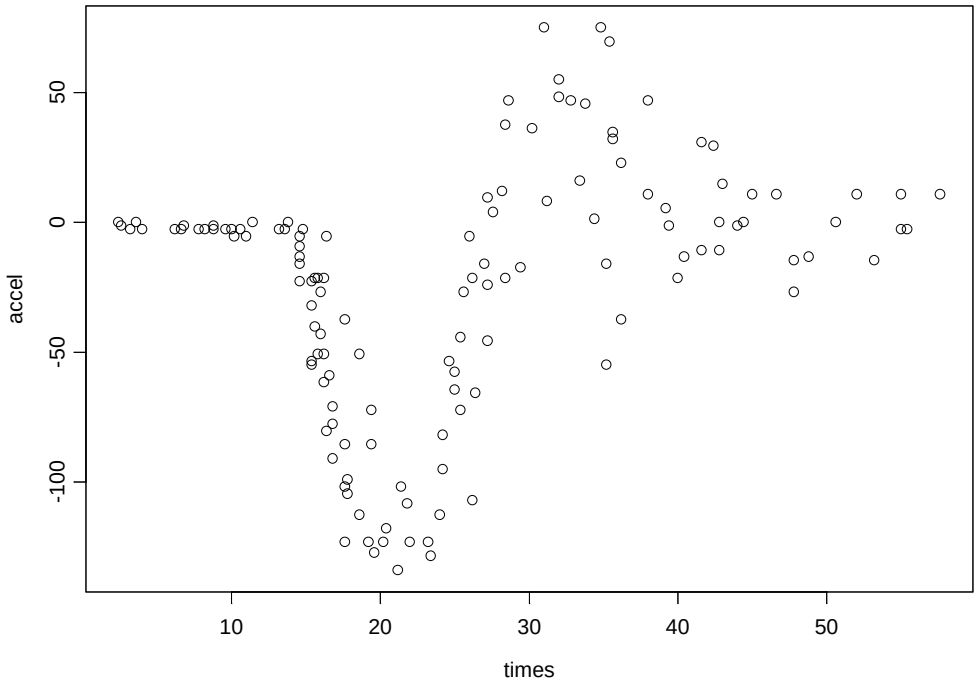
*Two quantitative variables*

- Decide which is *independent* and which is *dependent*
  - *outcome* is dependent
  - Other variable is independent
- Both variables are usually *quantitative*
- Useful tools: scatter plots, regression lines

*Example*

Head acceleration in simulated motorcycle accidents, used to test crash helmets.

- $n$=133 crashes (too many to list)
- `times:` in milliseconds after impact
  `accel:` in gee's




## Describing scatter plots

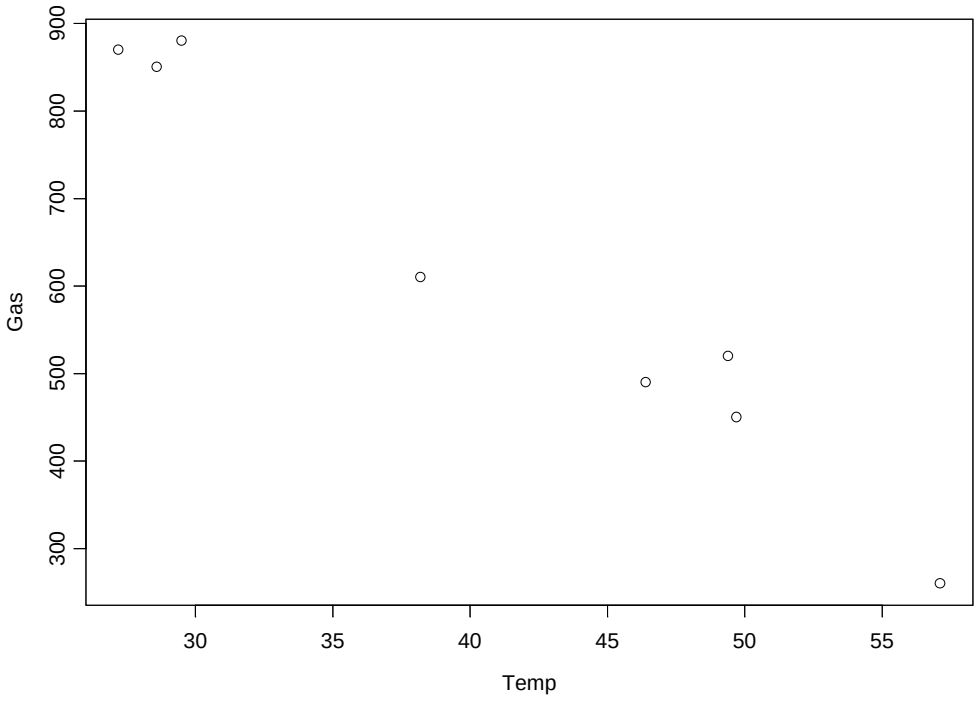
- Plotting, transforming if need be
- Relations between the variables
  - Positive (increasing) vs. negative (decreasing) association
  - linear, vs. nonlinear relationships
  - U-shaped relationships, vs. “no relationship”
- Clustering
- Outliers (in X or Y)
- Unequal Variability

*Example*

Natural Gas Consumption (Moore, p. 335).

- Average outside temperature (F)
- Average natural gas consumed (Cu Ft)

| Month | Oct | Nov | Dec | Jan | Feb | Mar | Apr | M |
|---|---|---|---|---|---|---|---|---|
| Temp | 49.4 | 38.2 | 27.2 | 28.6 | 29.5 | 46.4 | 49.7 | |
| Gas | 520 | 610 | 870 | 850 | 880 | 490 | 450 | 2 |

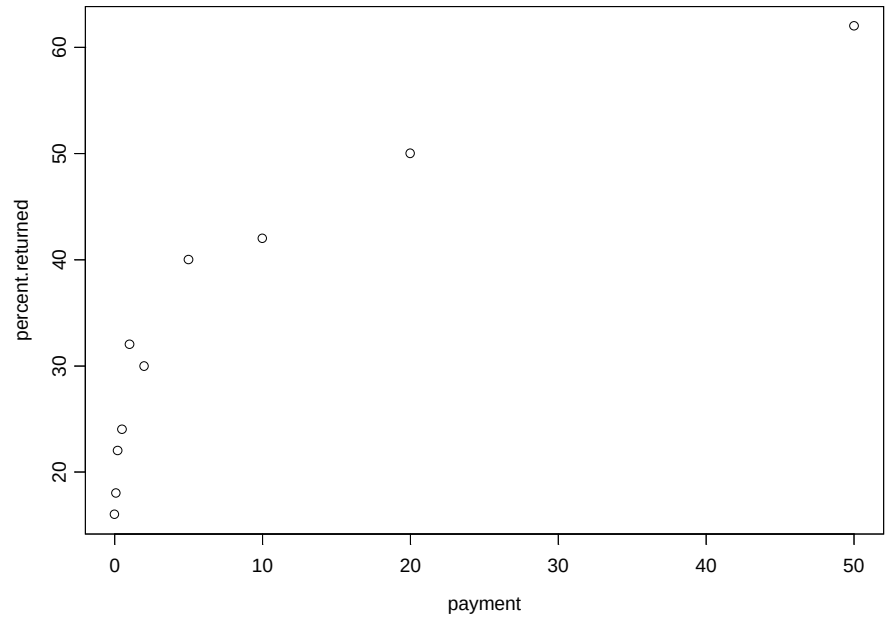



- Plotting, transforming if need be
- Relations between the variables
- Clustering
- Outliers (in X or Y)
- Unequal Variability

*Example*

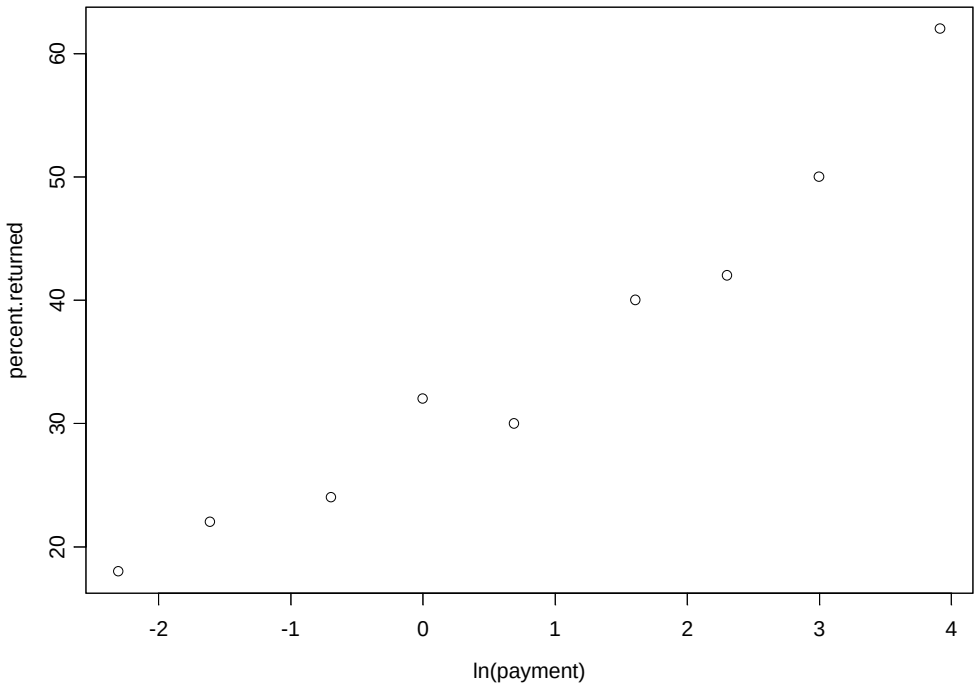
Results of a study of questionnaire response when payments are enclosed (Siegel and Morgan, p. 527).

| Pmt (Francs) | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| Pct Returned | 16 | 18.0 | 22.0 | 24.0 | 32 | 30 | 40 | 42 | 50 |




- Plotting, transforming if need be
- Relations between the variables
- Clustering
- Outliers (in X or Y)
- Unequal Variability

## *The effect of a transformation*




*Interpretation:*

- An increase of about 0.7 on the ln scale corresponds to doubling the orginal numbers.
- You have to double the payment to get each additional 5% of forms returned.