

# Interactive Data Graphics

- What is understood by the term interactive?
- What does interactive graphics mean?
- Interactive graphics
  - Principles
  - Practice
  - Particulars

# Interactivity in software

- PRIM9
- Excel
- Data Desk, JMP, ViSta
- ggobi
- R
- MANET, Mondrian, GAUGUIN, SEURAT
- iplots
- iPhone, iPad, ...

# Dynamic Graphics

(are ≠ interactive graphics)

- Changing point sizes
- Altering levels of alpha-blending
- Animation through categories
- Animated zooming (including censored zooming)
- Varying binwidths and other parameters
- Sliders
- Rotating plots
- ...
  - Movement not interaction

# Example: ratings (languageR)

Subjective frequency ratings, ratings of estimated weight, and ratings of estimated size, averaged over subjects, for 81 concrete English nouns.

A data frame with 81 observations on the following 14 variables.

<b>Word</b>	a factor with words as levels.
<b>Frequency</b>	a numeric vector of logarithmically transformed frequencies
<b>FamilySize</b>	a numeric vector of logarithmically transformed morphological family sizes
<b>SynsetCount</b>	a numeric vector with logarithmically transformed counts of the number of synonym sets in WordNet in which the word is listed
<b>Length</b>	a numeric vector for the length of the word in letters
<b>Class</b>	a factor with levels animal and plant
<b>FreqSingular</b>	a numeric vector for the frequency of the word in the singular
<b>FreqPlural</b>	a numeric vector with the frequency of the word in the plural
<b>DerivEntropy</b>	a numeric vector with the derivational entropies of the words
<b>Complex</b>	a factor coding morphological complexity with levels complex and simplex
<b>rInfl</b>	a numeric vector coding the log of ratio of singular to plural frequencies
<b>meanWeightRating</b>	a numeric vector for the estimated weight of the word's referent, averaged over subjects
<b>meanSizeRating</b>	a numeric vector for the estimated size of the word's referent, averaged over subjects
<b>meanFamiliarity</b>	a numeric vector with subjective frequency estimates, averaged over subjects

## R or Interactive Graphics

### R

#### Summary statistics

- means, sds, quantiles
- frequency tables
- correlations

#### Graphics

- single plots
- several plots one window
- small multiples

#### Models

### Mondrian

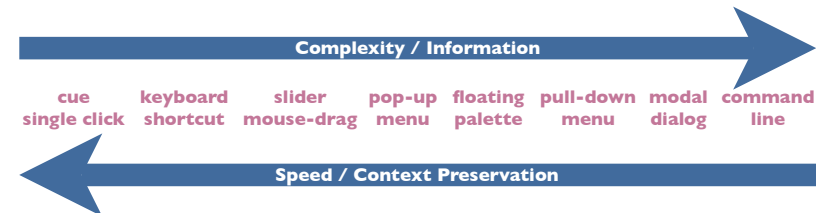
#### Graphics

- multiple plots
- multivariate plots

#### Interaction

- querying
- linking
- zooming
- varying
- reordering

## Interactivity hierarchy



**Fig. 4.11.** Hierarchy of user interface controls according to their information content, complexity, speed, and context preservation.

Ideal is direct manipulation of the data objects and statistical objects with immediate response

## IG Advantages

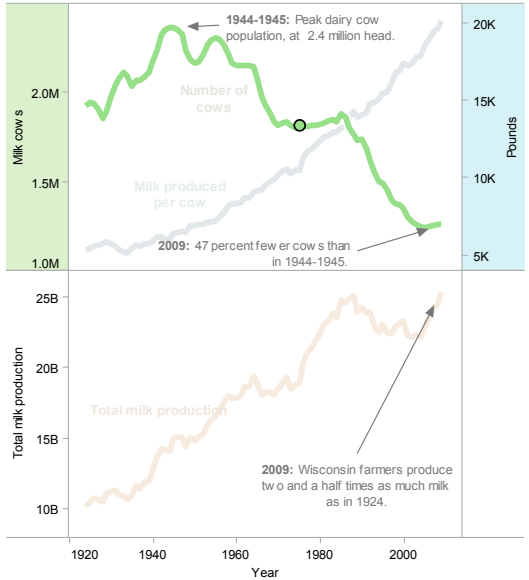
- Direct querying
- Multivariate information via linking
- Fast, flexible analyses (including sensitivity analyses)
- Running through alternatives quickly
- Experimental reformatting
- Versatile reordering
- Generate ideas/hypotheses
- and
  - not letting computing get in the way of thinking

## IG Disadvantages

- Not mathematically defined
- Difficult to record the process
- Cannot replicate analyses
- Difficult to save results of analyses
- Can often not test results statistically
- Not presentation graphics quality
- Data dredging: you always find something

# Fewer cows, but more milk

## Dairy cows and milk production, 1924-2009



- Legend**
- Milk cows
  - Milk per cow
  - Milk production

**Top chart:** Recent Wisconsin dairy cow populations are at the lowest levels in 85 years -- but today's cow produces three times as much milk as her 1924 counterpart. **Lower chart:** Since then, total dairy production has increased by two and half times.

Data source: National Agricultural Statistics Service.  
 Chart: Kate Golden/  
 WisconsinWatch.org.