

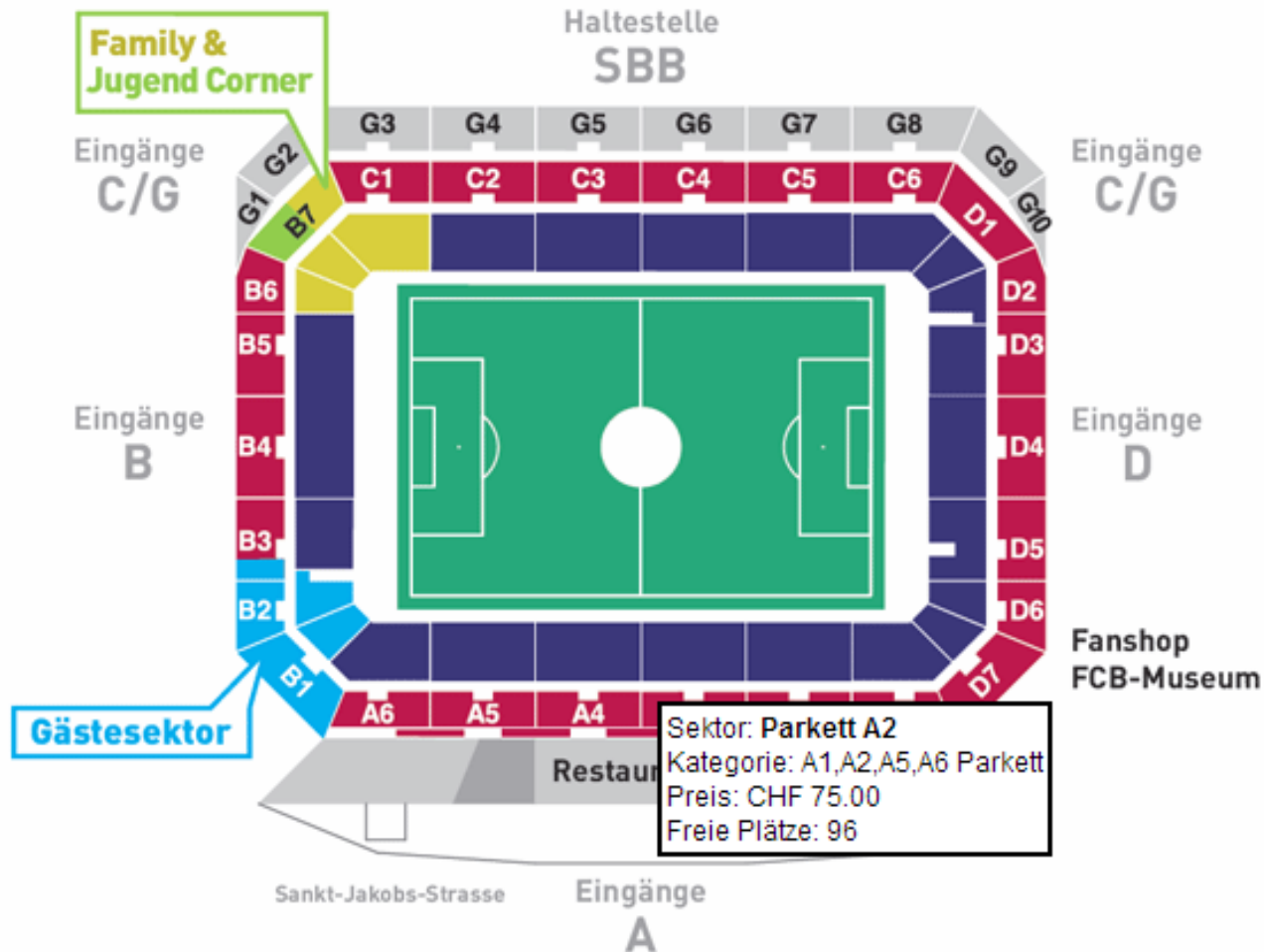
# GRAPHICS OF CLINICAL DATA: THE GOOD, THE BAD, AND THE UGLY

PHUSE SINGLE DAY CONFERENCE ON DATA VISUALIZATION.  
BASEL, SWITZERLAND  
JULY 3, 2014

Andreas Krause  
Lead Scientist Modeling and Simulation  
Department of Clinical Pharmacology  
Actelion Pharmaceuticals, Allschwil, Switzerland

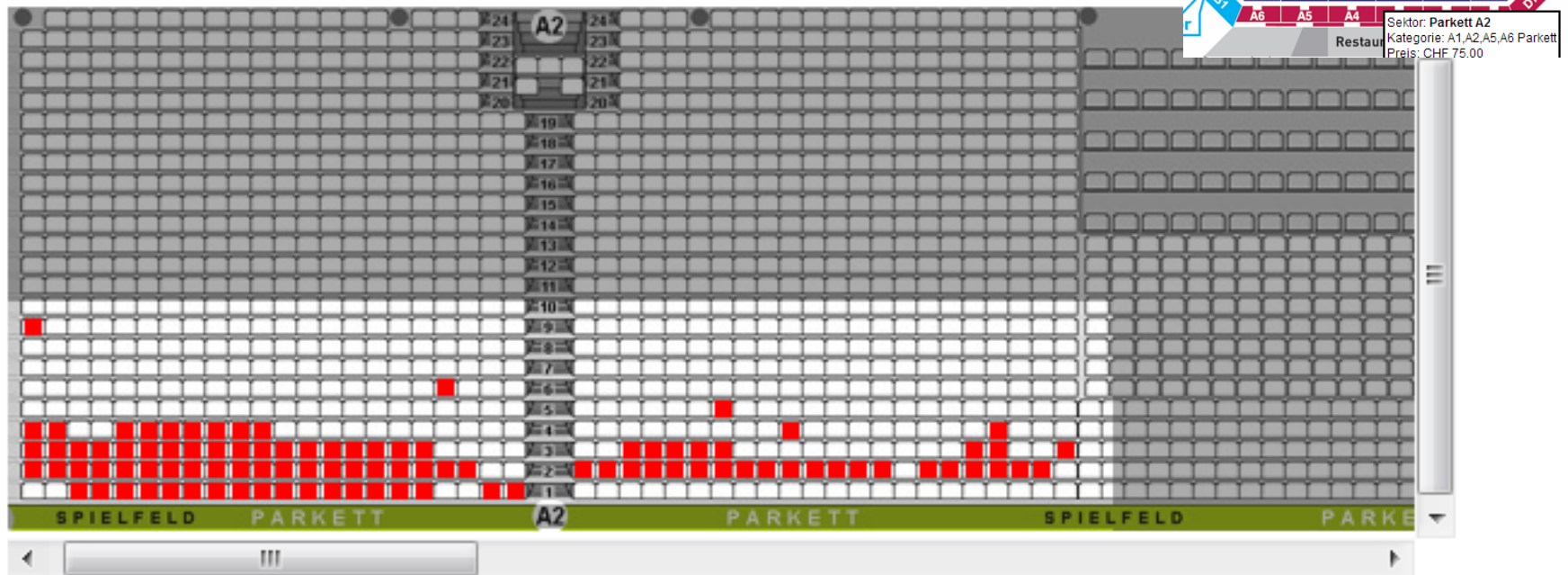
# CASE STUDY: PURCHASING A STADIUM TICKET

Bitte wählen Sie den gewünschten Sektor:



# CASE STUDY: PURCHASING A STADIUM TICKET

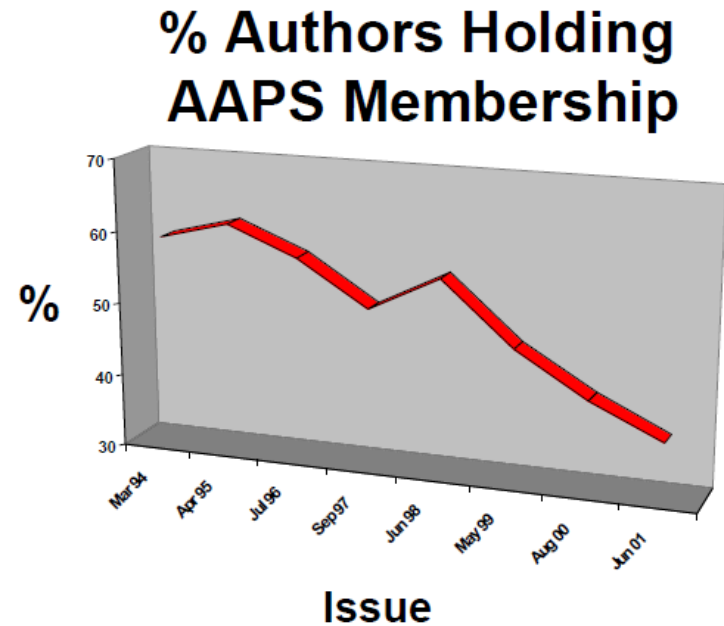
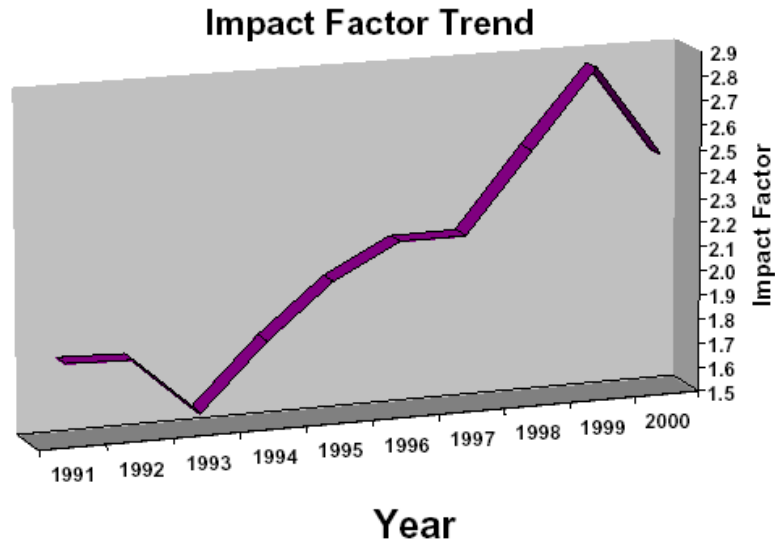
Bitte wählen Sie die gewünschten Plätze:



The seats marked **red** are the available seats!!!

# HIGHER DIMENSIONS: REAL LIFE EXAMPLES

---



- ▶ Top left: more dramatic increase by tilting the x-axis upwards and looking at the curve from the bottom.
- ▶ Bottom right: same trick for a downward trend!

# THE GOAL

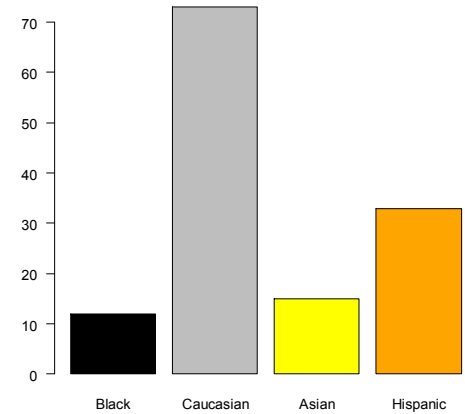
---

- ▶ Communicate the information in the data
  - The real story, no fudging
  - Marketing Graph - A pictorial representation which uses three dimensions, four colours and five cartoons to show one fact which probably isn't true.
    - From: The Devil's Drug Development Dictionaries. <http://www.senns.demon.co.uk/wdict.html>

# PRINCIPLE: THE FOCUS IS ON THE DATA

---

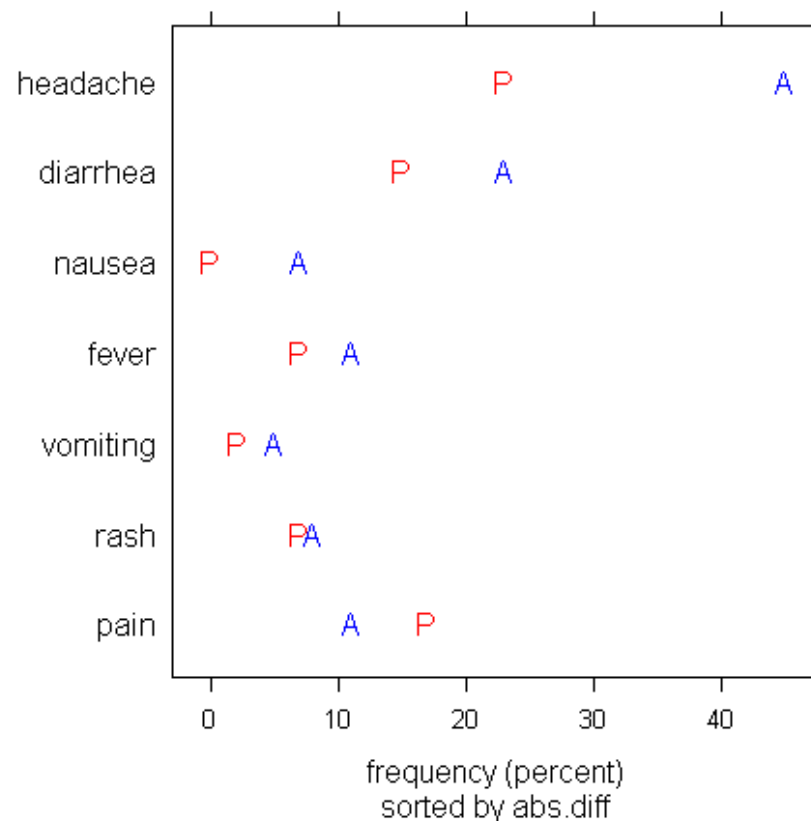
- ▶ Maximize the “data to ink ratio” (Tufte)
- ▶ Clear purpose
- ▶ Encourage and facilitate visual comparison
  - Use few different plot types
  - Add (faint) horizontal and/ or vertical lines
  - Choose appropriate order of categories for categorical data
- ▶ Color and gray scale use only if it supports the analytical process
- ▶ **Easy:** Positioning/ comparison along a common scale, relative lengths
- ▶ **Difficult:**
  - Comparison of angles and slopes (pie charts)
  - Area and volume (bar charts, 3D)
  - Color, shading, saturation: discerning ordering



# A PICTURE IS WORTH A THOUSAND TABLES

| Event    | Cases on active | Cases on placebo | Relative Diff. |
|----------|-----------------|------------------|----------------|
| headache | 45              | 23               | 96%            |
| diarrhea | 23              | 15               | 53%            |
| pain     | 11              | 17               | -35%           |
| nausea   | 7               | 0                | -              |
| vomiting | 5               | 2                | 150%           |
| fever    | 11              | 7                | 57%            |
| rash     | 8               | 7                | 14%            |

- ▶ Safety Data sorted by difference in AE rate
- ▶ Easy reading
  - Highest incidence rate
  - Lowest incidence rate
  - Diff. btw active and placebo
  - Comparison between AEs
- ▶ Intuitive symbols(P, A)

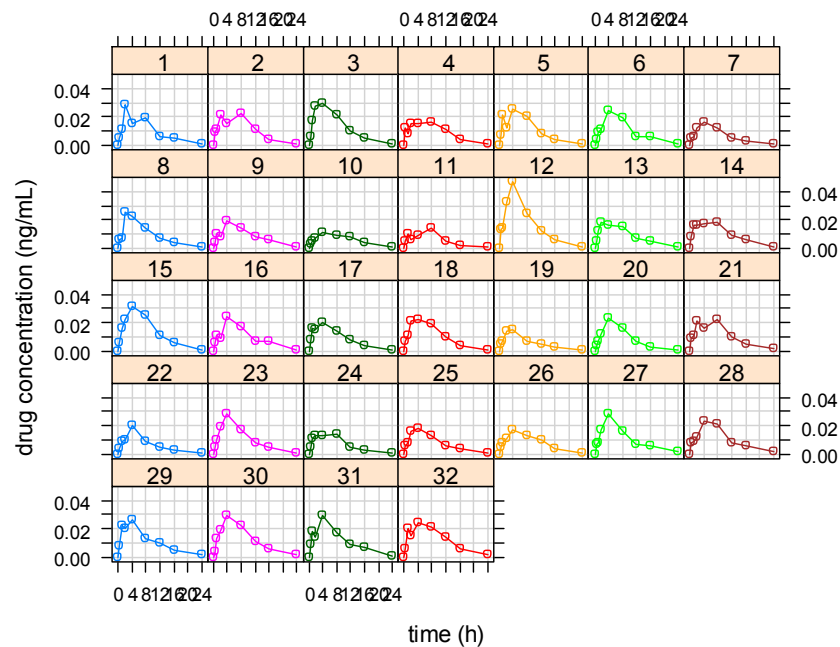
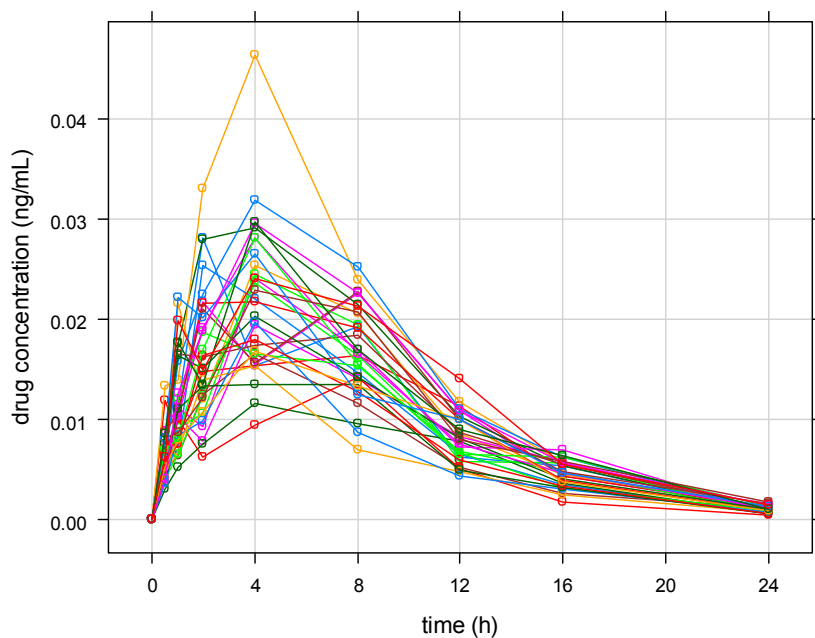


# PRINCIPLES: SOME CONSEQUENCES

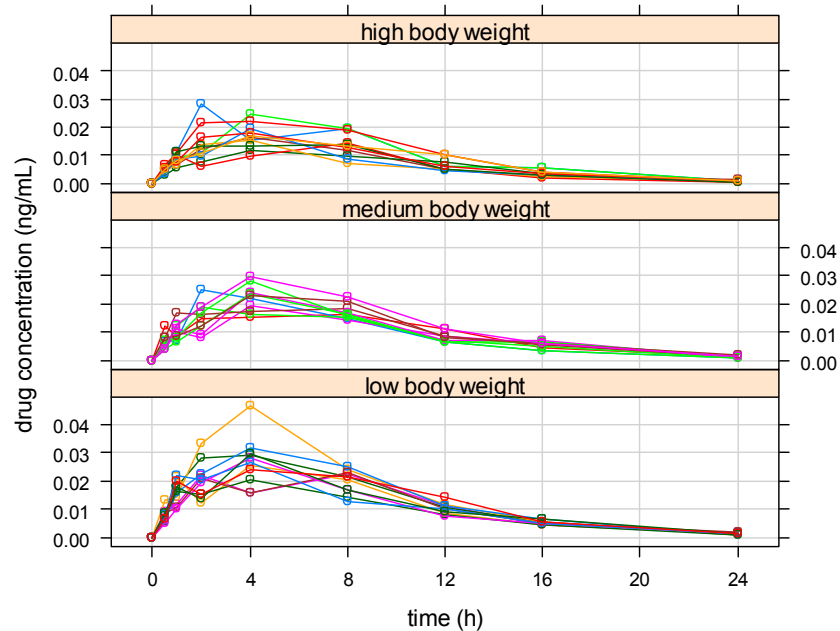
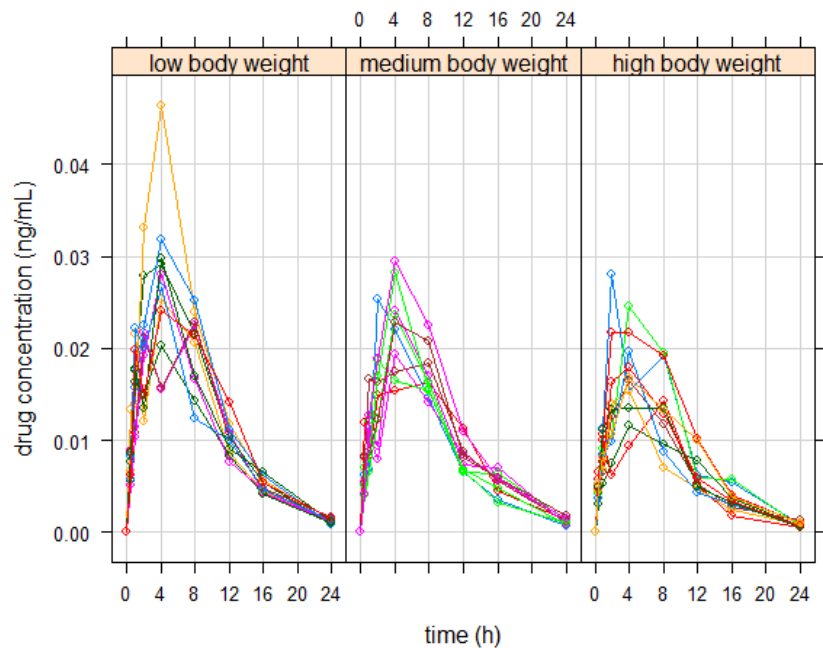
---

- ▶ Aggregation
  - If feasible, show all data
    - Aggregation can be misleading (e.g., choice of categories)
    - Aggregate to an appropriate level
  - If not, don't: A scatterplot with 10,000 dots does not show anything
  
- ▶ Graph elements
  - Keep the axes out of the graphs (tickmarks, labels, everything)
  - No legend inside graphs
  - Annotation only if necessary
  - Color only if it helps visualization
  - Make reading of the graph intuitive
  
- ▶ **Principle: Think before you graph!**

# PRINCIPLE: ENCOURAGE COMPARISON

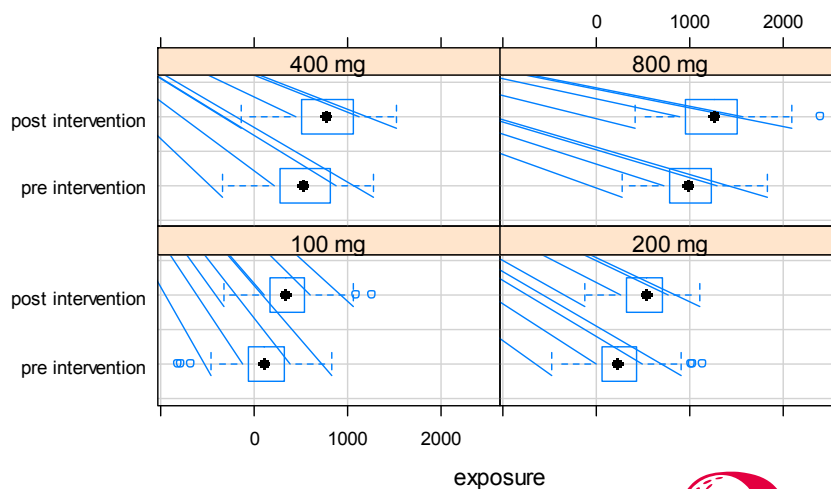
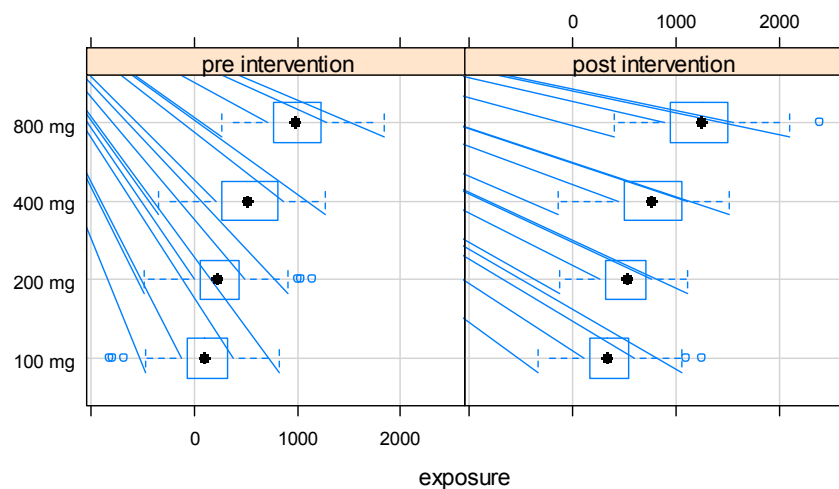


# PRINCIPLE: ENCOURAGE COMPARISON



# PRINCIPLE: ENCOURAGE COMPARISON

- ▶ AUC pre and post an event for four doses
- ▶ Left: compare exposure/ dose by pre/ post
- ▶ Right: compare pre and post by dose
- ▶ Same eight box plots, different arrangement! To answer different questions



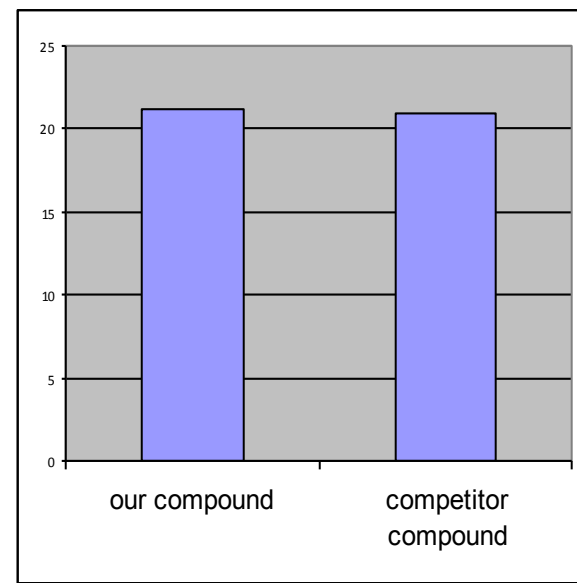
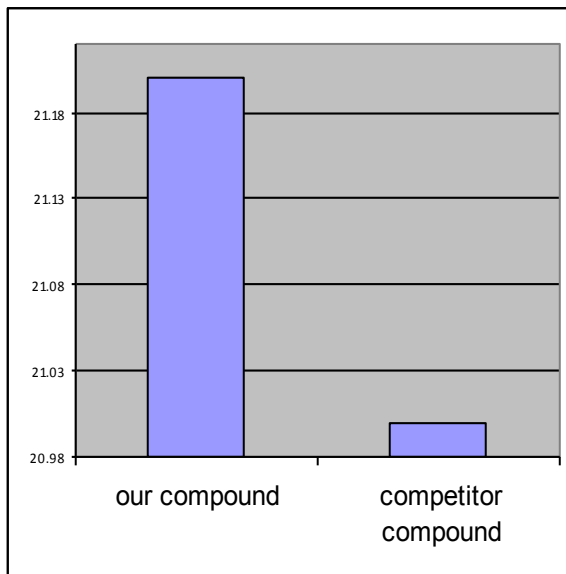
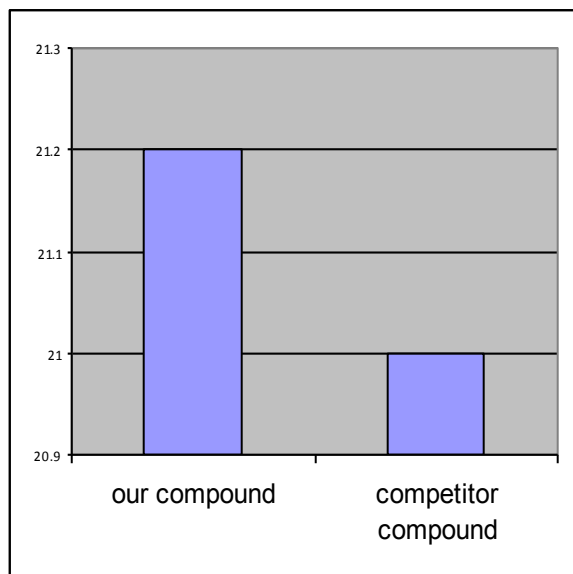
# ELEMENTS OF GRAPHICS

---

- ▶ Axes
- ▶ Lines
- ▶ Symbols
- ▶ Colors
- ▶ Legends

# AXIS RANGES

- ▶ Choosing the range can help or destroy pattern recognition
- ▶ The graphs below just differ by the choice of the y-axis range
- ▶ Consider inclusion of the origin



# CHOOSING THE AXES

---

Intuition: the y-variable is dependent on the x-variable:

y is modeled or *graphed* as a function of x

Examples:

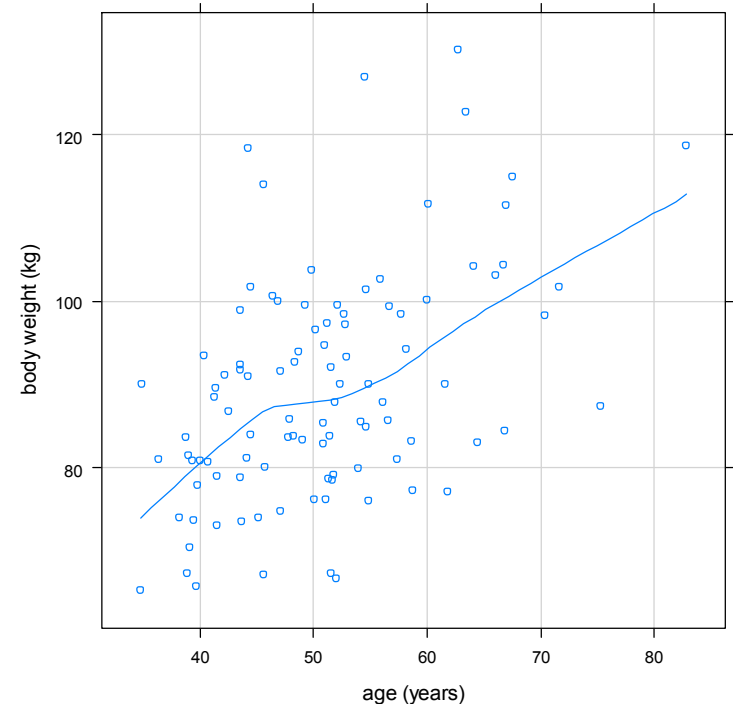
Heart rate is a function of the body weight

Body weight is a function of gender and age

Fitted versus observed values

Observed  $\sim N(m, s \mid \text{model})$

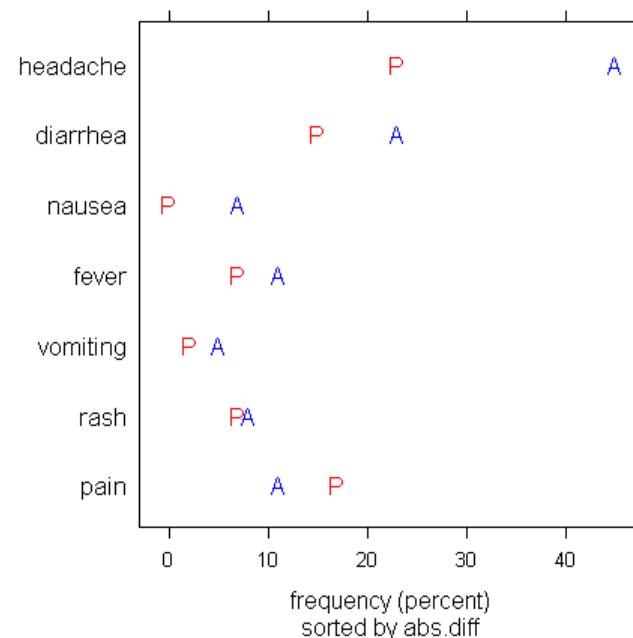
(*not vice versa*)



# POINT SYMBOLS

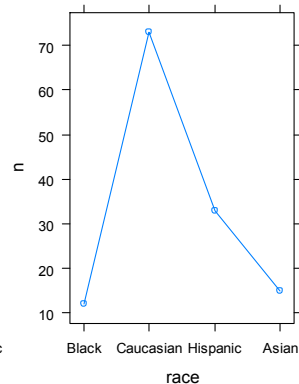
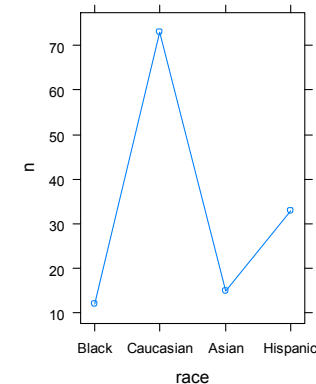
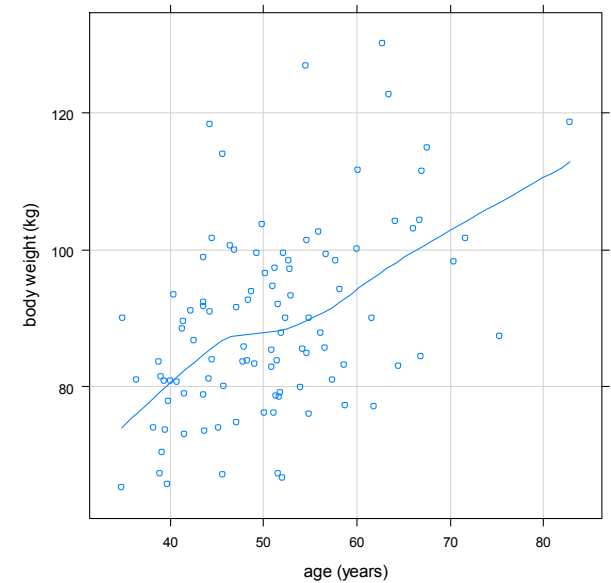
- ▶ Open circles are easier to discern than filled bullets
- ▶ Single pixels are too small
- ▶ Circles, triangles, and other symbols
  - Require a legend
  - Force to switch between graph and legend
  - Are difficult to identify
- ▶ Symbols or colors: consider using a monotonicity
  - 2, 3, 4, 5 edges (line, triangle, square, ...) to represent increasing doses, ages, etc.
  - Colors from light (low dose) to dark (high dose)
- ▶ Symbols can be meaningful
  - “A” for active, “L” for low, “H” for high dose
  - “0”, “2”, “5”, “8” for 0, 20, 50, 80 mg

| event    | active | placebo | relative difference | p-value (etc.) |
|----------|--------|---------|---------------------|----------------|
| headache | 45     | 23      | +96%                |                |
| diarrhea | 23     | 15      | +53%                |                |
| pain     | 11     | 17      | -35%                |                |
| nausea   | 7      | 0       | -                   |                |
| vomiting | 5      | 2       | +150%               |                |
| fever    | 11     | 7       | +57%                |                |
| rash     | 8      | 7       | +14%                |                |



# LINES

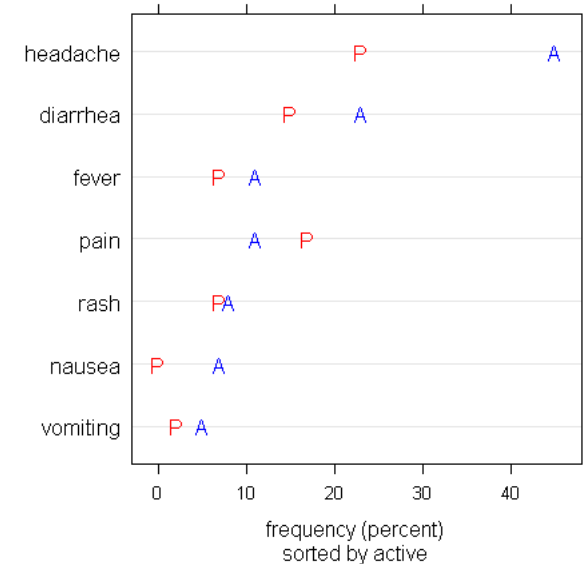
- ▶ A line suggests an order
  - A trend line (in time)
  - A regression line: low x-values to high x-values
- ▶ A line is useless if not irritating if there is no order
  - Connecting races or genders
  - Different order even suggests different interpretation!
- ▶ Use monotonicity to represent order:
  - Examples: doses, ages, body weights
    - Line widths: thin, medium, thick
    - Colors light gray, dark gray, black
    - Dotted and dashed lines



# LEGENDS

---

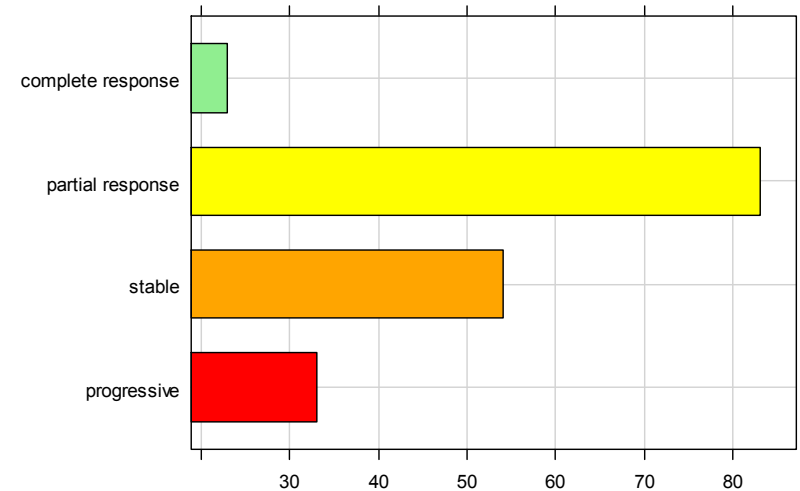
- ▶ “If it needs a legend, you might want to think again”
- ▶ Introduce an order that maps an order in the data
  - Green/ yellow/ red
  - Dark green/ light green/ yellow/ orange/ light red/ dark red
  - From dark to light
  - From thick to thin lines
- ▶ Introduce a logic such that looking at the legend *once* suffices
  - “L”, “M”, “H” for low, medium and high dose
  - “P” for placebo and “A” for active



# COLORS

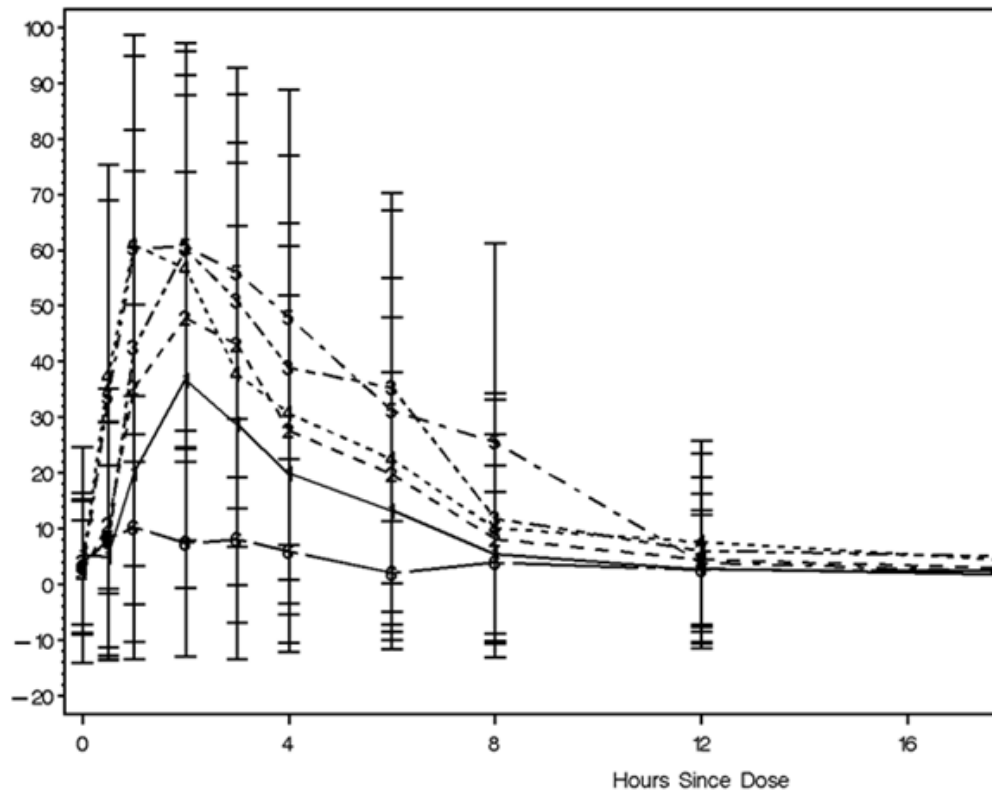
---

- ▶ Use intuitive colors
  - Connotation
    - Green: “good”
    - Red: “bad” (red alert)
    - Yellow: “watch out” (in between green and red)
- ▶ Example: patient response to treatment
  - Disease improvement: green
  - Disease worsening: red
  - Unchanged: yellow



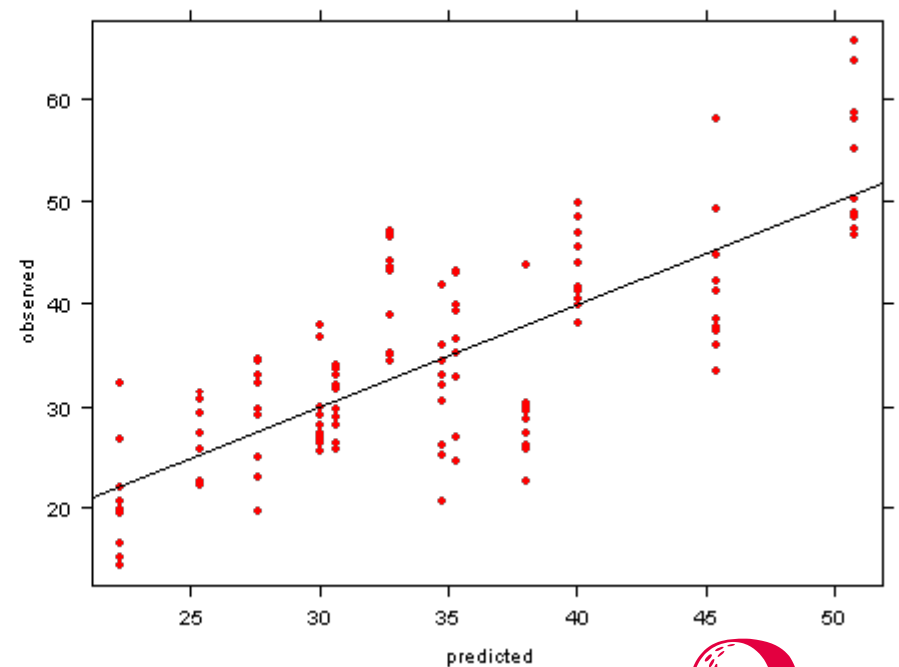
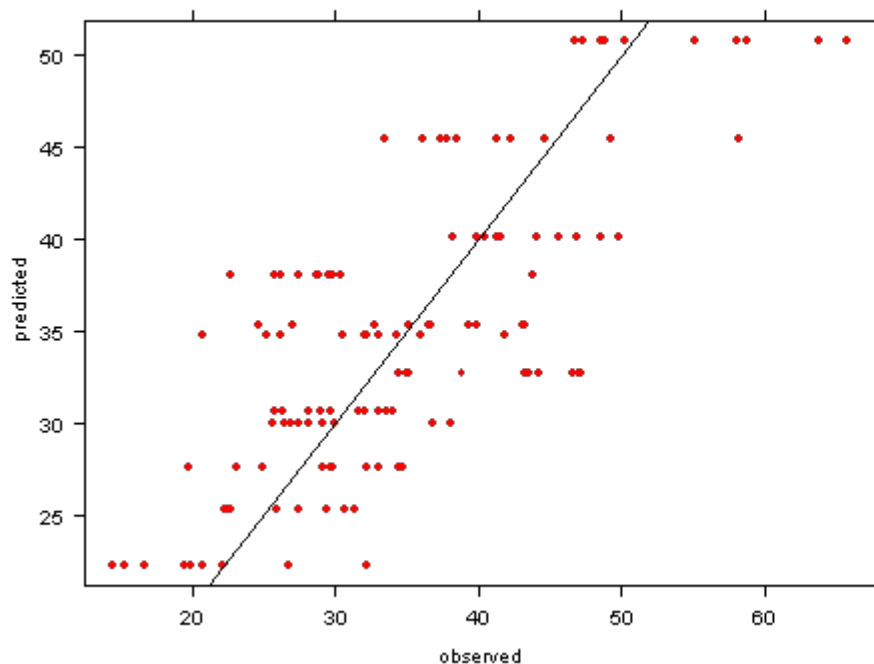
# PARTICULAR APPLICATIONS

---



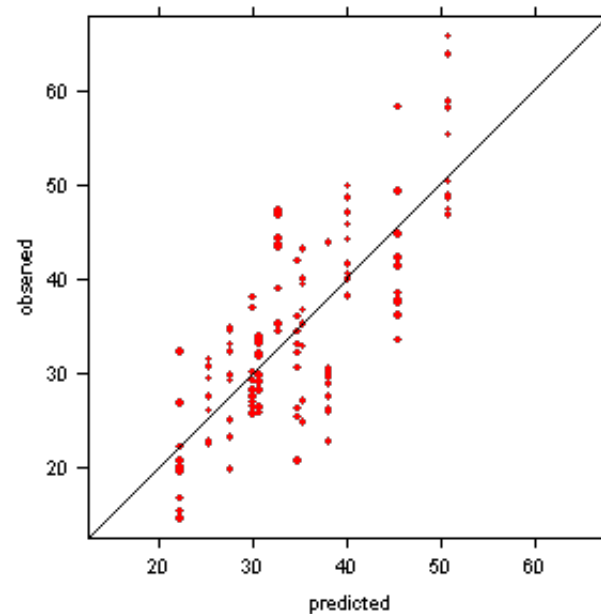
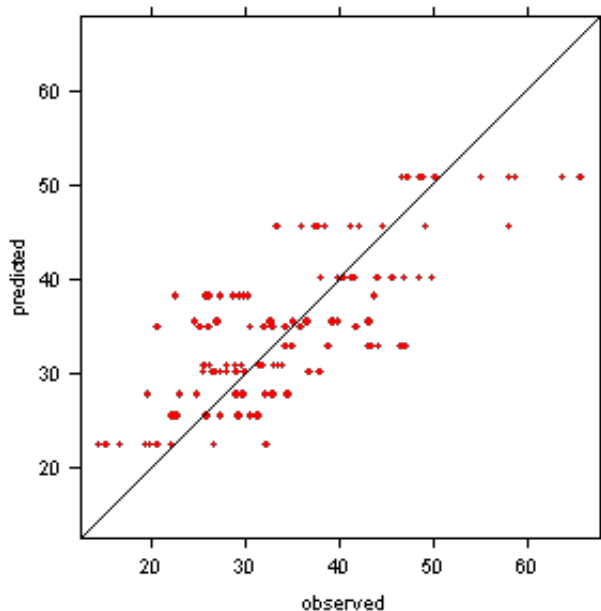
# COMPARING LIKE WITH LIKE

- ▶ Two comparisons: Which set of (x, y) values is more similar?
- ▶ The two data sets are the same!
  - Just that x and y are swapped
- ▶ So why is the visual perception so different?



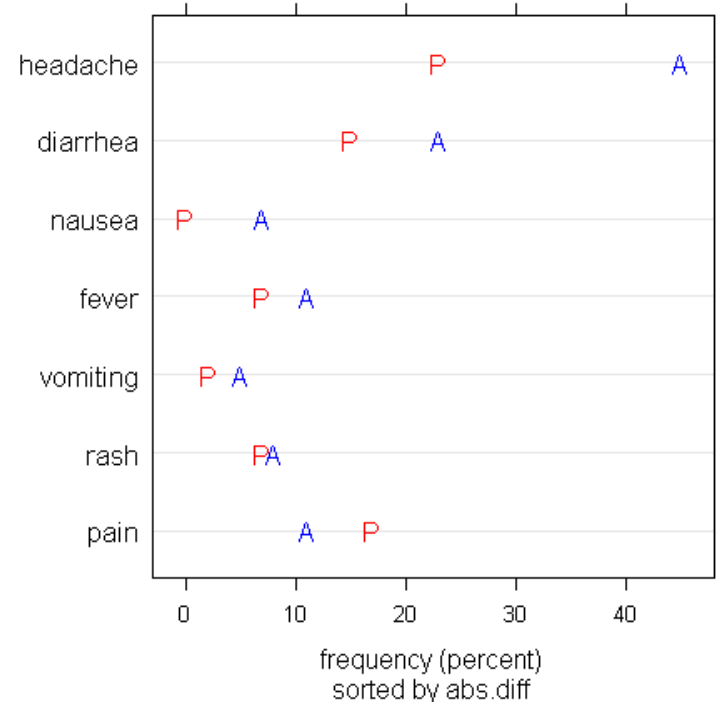
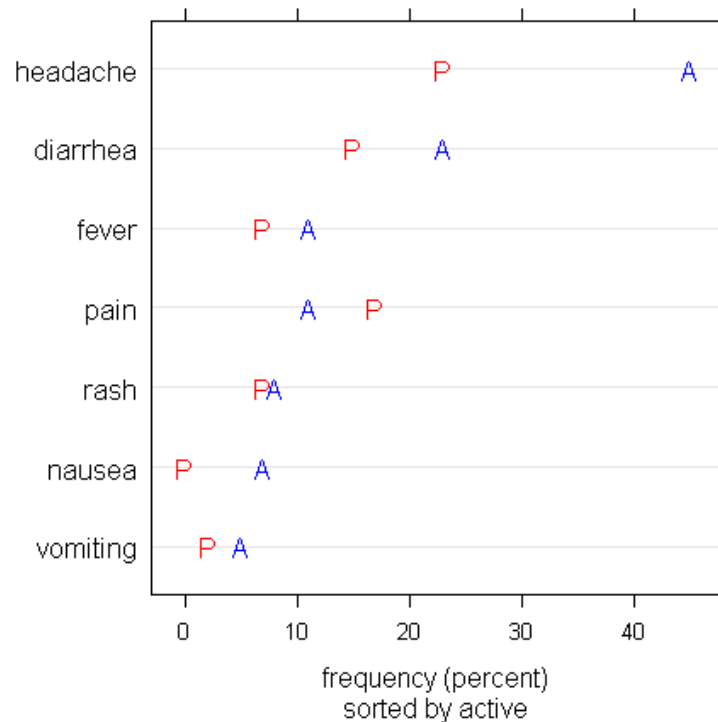
# COMPARING LIKE WITH LIKE

- ▶ Same axis ranges
- ▶ One unit corresponds to the same number of pixels on both axes
  - The graphs are square
  - Consequence: the identity line has a 45 degree slope
- ▶ Avoids visual bias



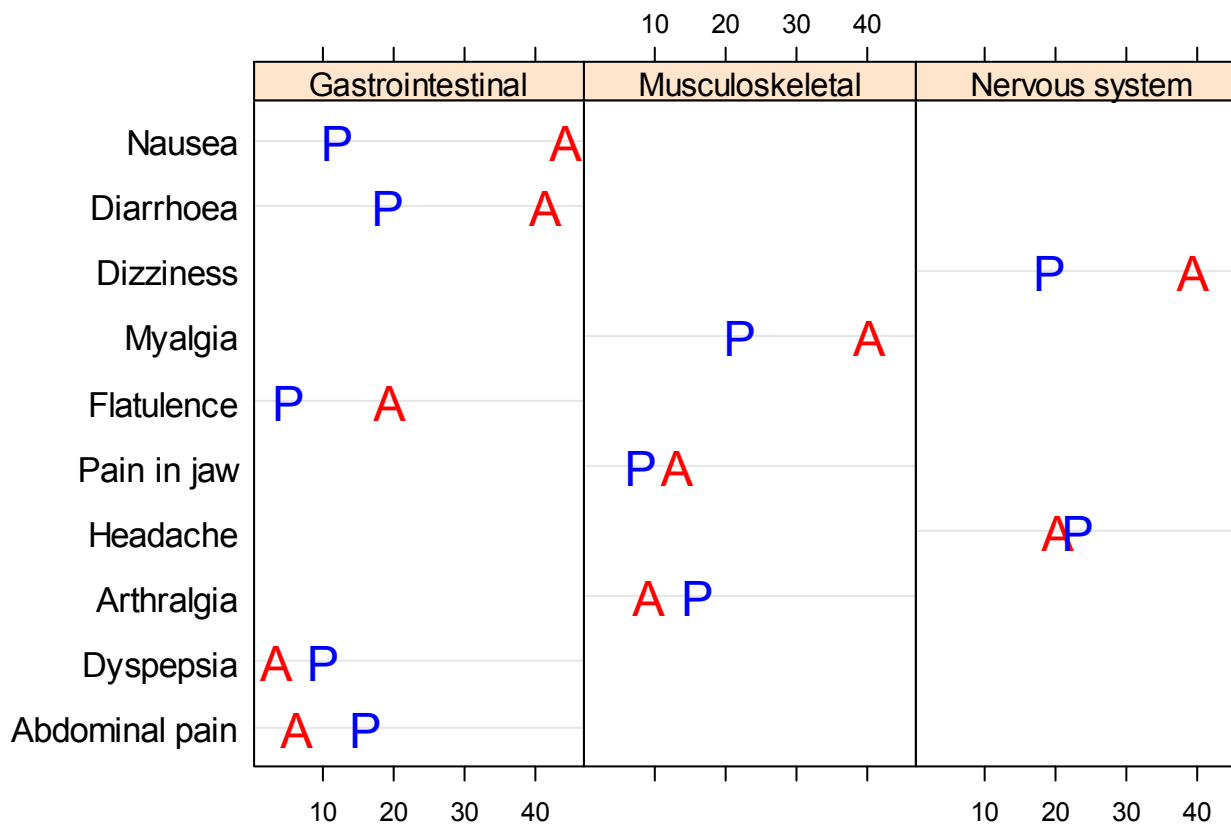
# COMPARISON BY ORDERING

- ▶ To facilitate reading of graphs, consider the ordering of categories.
- ▶ Left: categories ordered by percentage occurrence in active treatment
- ▶ Right: categories ordered by absolute difference btw active and placebo



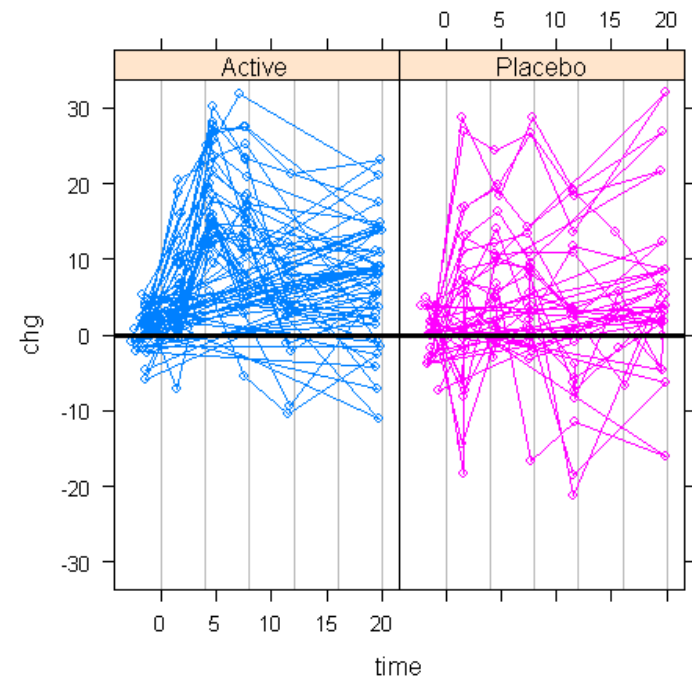
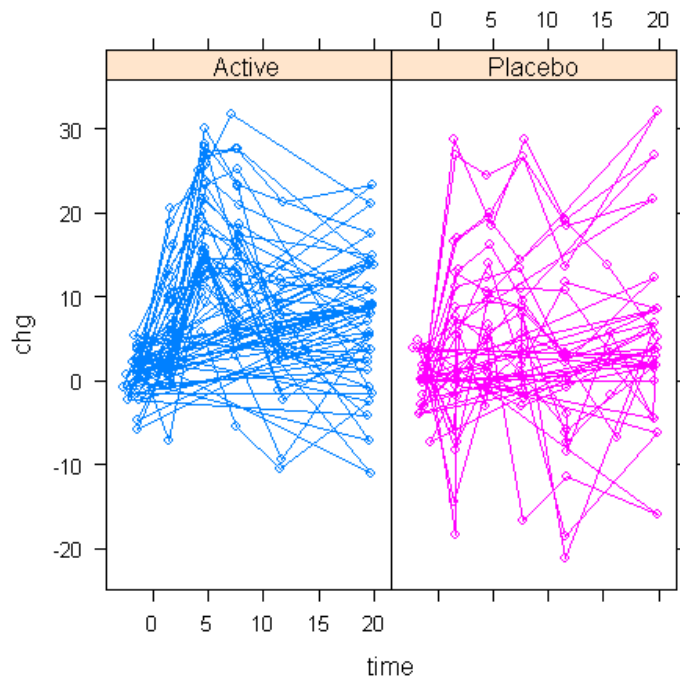
# COMPARISON BY ORDERING AND GROUPING

- Group AEs into categories (nervous system, respiratory, skin, etc.)



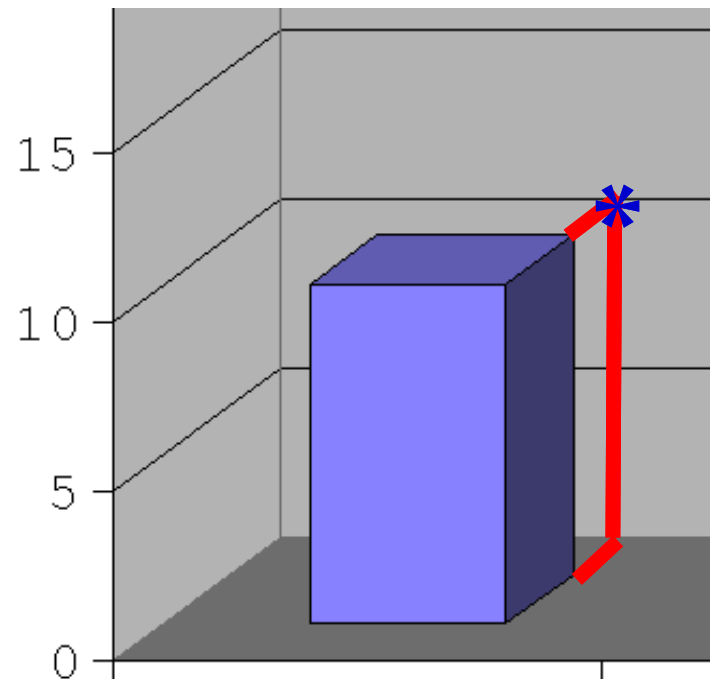
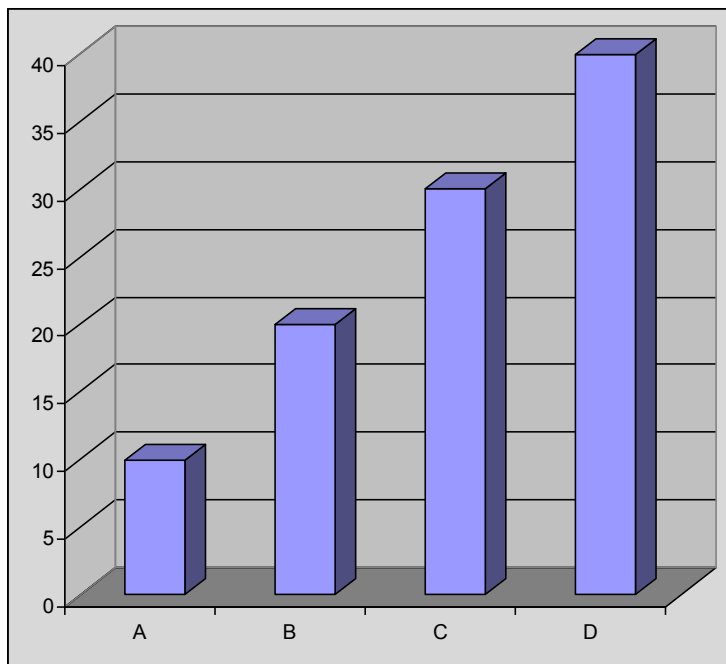
# CHANGE FROM BASELINE

- ▶ Change from baseline = change from 100% or 1
- ▶ To avoid misleading visual perception, consider
  - a graph symmetric around “no change”
  - Addition of a supportive line of no change



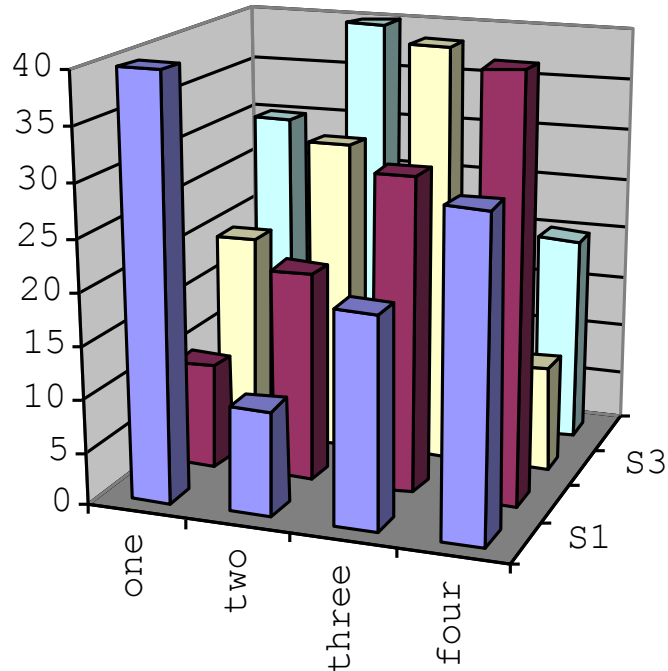
# HIGHER DIMENSIONS

- ▶ A standard 3D graph (the default)
- ▶ Exercise: Read off the values.

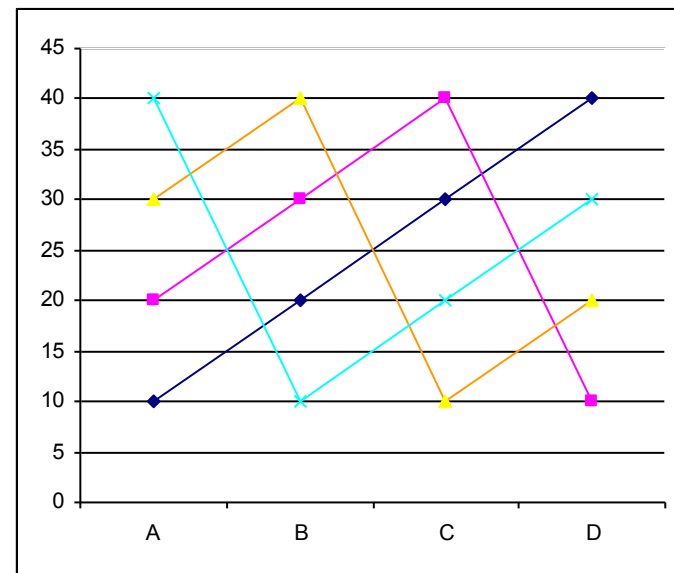


# HIGHER DIMENSIONS

- ▶ A 3D data set on a tilted surface
- ▶ What are the values? What bar is the highest?
- ▶ Other aspects: angles and order change perception
- ▶ Alternative: four lines with four points each: Might not look as “fancy”.

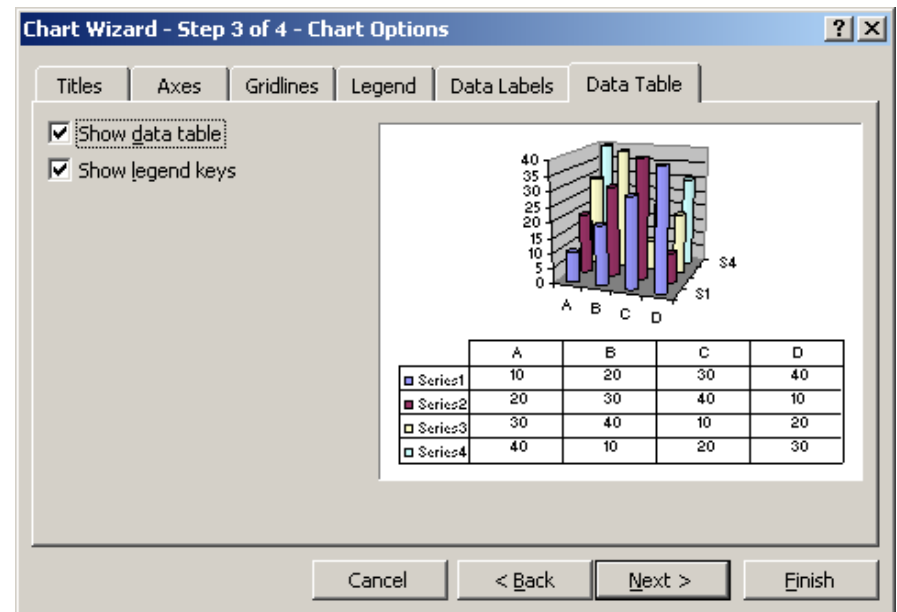
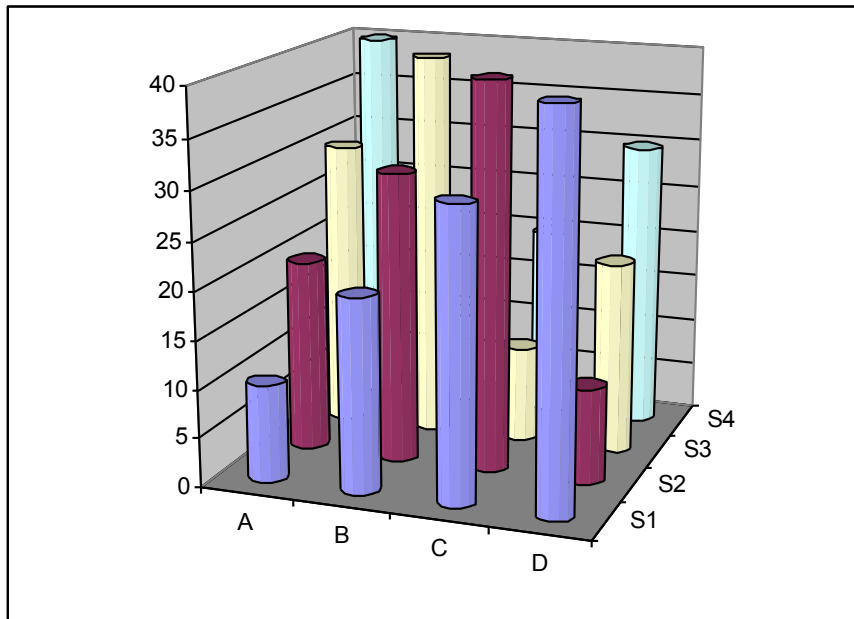


Might be more useful though



## HIGHER DIMENSIONS: ADD A TABLE?

- ▶ Optionally, you can show a table
- ▶ Well...
  - Why show a table if the graph is useful?

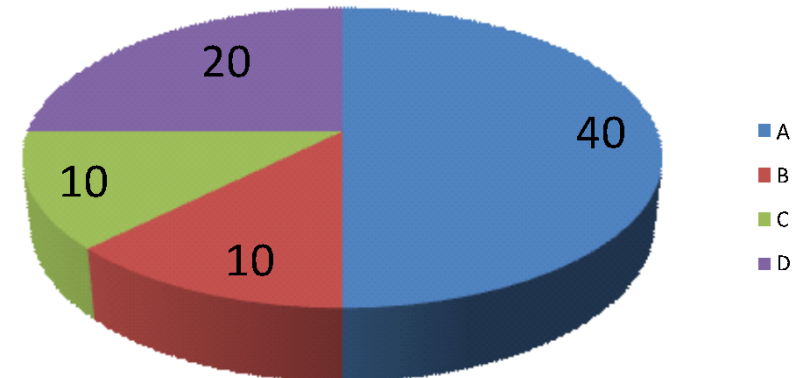
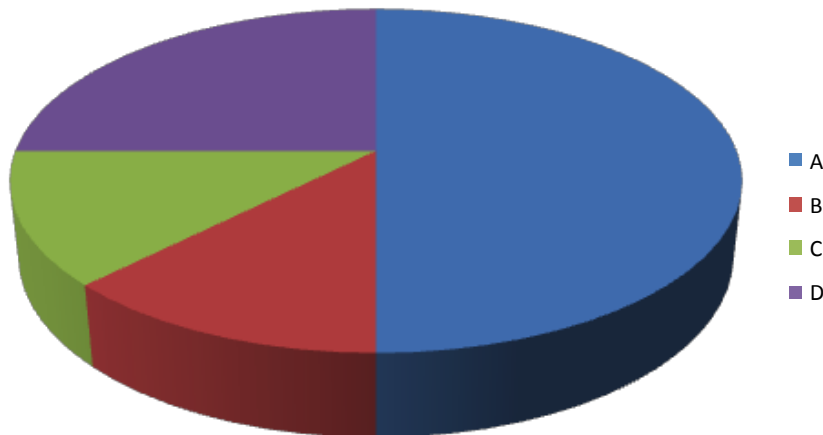


# HIGHER DIMENSIONS: PIE CHARTS

---

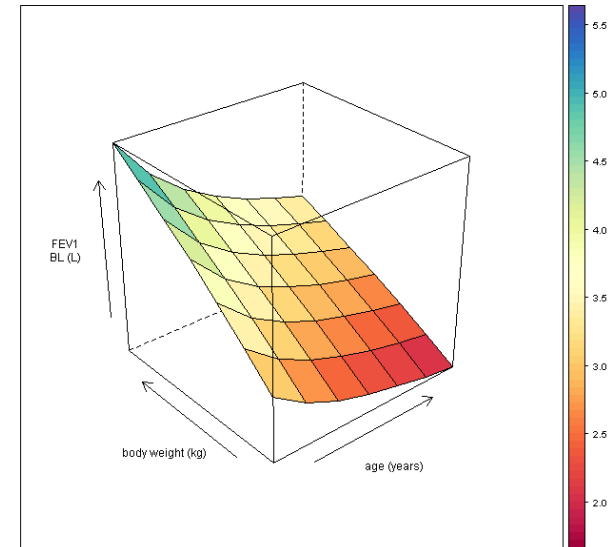
- ▶ The yellow area has many more pixels due to the shaded yellow side of the pie in front
  - Values: yellow: 20, green: 30
- ▶ One of the first options: add the numbers
- ▶ But what's the point of the graph then?
- ▶ Use a table to show the 4 numbers

|   |    |
|---|----|
| A | 40 |
| B | 10 |
| C | 10 |
| D | 20 |

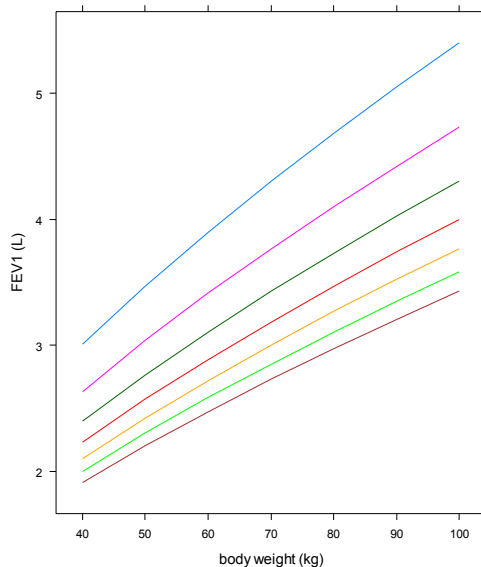


# HIGHER DIMENSIONS: SLICING THE SURFACE

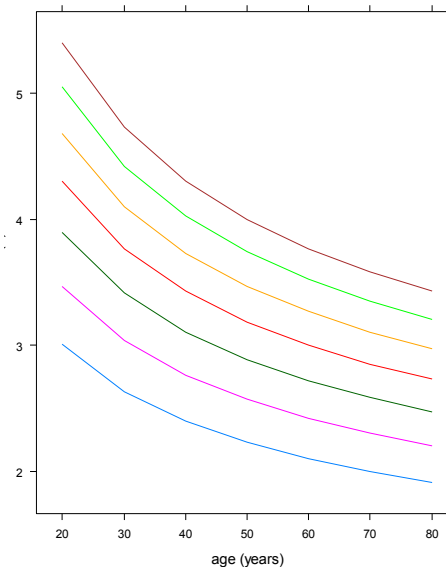
- ▶ Illustrate relationships between variables
  - FEV1 as a function of age and body weight
  - FEV1 increases with body weight
  - FEV1 decreases with age
- ▶ Reading off values is difficult (y-axis and/or color legend here). Drop axes!
- ▶ Use 2D to read off values



FEV1 baseline (L) estimated  
age 20 to 80 years; body weight 40 to 100 kg



age categories  
20  
30  
40  
50  
60  
70  
80



body weight categories  
40  
50  
60  
70  
80  
90  
100

# INSPIRATION: EXAMPLE APPLICATIONS COMBINING THE PRINCIPLES

---

# TABLES AND GRAPHS (1)

| Cancer site                      | Relative survival rate, % (SE) |            |            |            |
|----------------------------------|--------------------------------|------------|------------|------------|
|                                  | 5 years                        | 10 years   | 15 years   | 20 years   |
| Oral cavity and pharynx          | 56.7 (1.3)                     | 44.2 (1.4) | 37.5 (1.6) | 33.0 (1.8) |
| Oesophagus                       | 14.2 (1.4)                     | 7.9 (1.3)  | 7.7 (1.6)  | 5.4 (2.0)  |
| Stomach                          | 23.8 (1.3)                     | 19.4 (1.4) | 19.0 (1.7) | 14.9 (1.9) |
| Colon                            | 61.7 (0.8)                     | 55.4 (1.0) | 53.9 (1.2) | 52.3 (1.6) |
| Rectum                           | 62.6 (1.2)                     | 55.2 (1.4) | 51.8 (1.8) | 49.2 (2.3) |
| Liver and intrahepatic bile duct | 7.5 (1.1)                      | 5.8 (1.2)  | 6.3 (1.5)  | 7.6 (2.0)  |
| Pancreas                         | 4.0 (0.5)                      | 3.0 (0.5)  | 2.7 (0.6)  | 2.7 (0.8)  |
| Larynx                           | 68.8 (2.1)                     | 56.7 (2.5) | 45.8 (2.8) | 37.8 (3.1) |
| Lung and bronchus                | 15.0 (0.4)                     | 10.6 (0.4) | 8.1 (0.4)  | 6.5 (0.4)  |
| Melanomas                        | 89.0 (0.8)                     | 86.7 (1.1) | 83.5 (1.5) | 82.8 (1.9) |
| Breast                           | 86.4 (0.4)                     | 78.3 (0.6) | 71.3 (0.7) | 65.0 (1.0) |
| Cervix uteri                     | 70.5 (1.6)                     | 64.1 (1.8) | 62.8 (2.1) | 60.0 (2.4) |
| Corpus uteri and uterus, NOS     | 84.3 (1.0)                     | 83.2 (1.3) | 80.8 (1.7) | 79.2 (2.0) |
| Ovary                            | 55.0 (1.3)                     | 49.3 (1.6) | 49.9 (1.9) | 49.6 (2.4) |
| Prostate                         | 98.8 (0.4)                     | 95.2 (0.9) | 87.1 (1.7) | 81.1 (3.0) |
| Testis                           | 94.7 (1.1)                     | 94.0 (1.3) | 91.1 (1.8) | 88.2 (2.3) |
| Urinary bladder                  | 82.1 (1.0)                     | 76.2 (1.4) | 70.3 (1.9) | 67.9 (2.4) |
| Kidney and renal pelvis          | 61.8 (1.3)                     | 54.4 (1.6) | 49.8 (2.0) | 47.3 (2.6) |
| Brain and other nervous system   | 32.0 (1.4)                     | 29.2 (1.5) | 27.6 (1.6) | 26.1 (1.9) |
| Thyroid                          | 96.0 (0.8)                     | 95.8 (1.2) | 94.0 (1.6) | 95.4 (2.1) |
| Hodgkin's disease                | 85.1 (1.7)                     | 79.8 (2.0) | 73.8 (2.4) | 67.1 (2.8) |
| Non-Hodgkin lymphomas            | 57.8 (1.0)                     | 46.3 (1.2) | 38.3 (1.4) | 34.3 (1.7) |
| Multiple myeloma                 | 29.5 (1.6)                     | 12.7 (1.5) | 7.0 (1.3)  | 4.8 (1.5)  |
| Leukaemias                       | 42.5 (1.2)                     | 32.4 (1.3) | 29.7 (1.5) | 26.2 (1.7) |

Rates derived from SEER 1973-98 database (both sexes, all ethnic groups).<sup>12</sup>  
NOS=not otherwise specified.

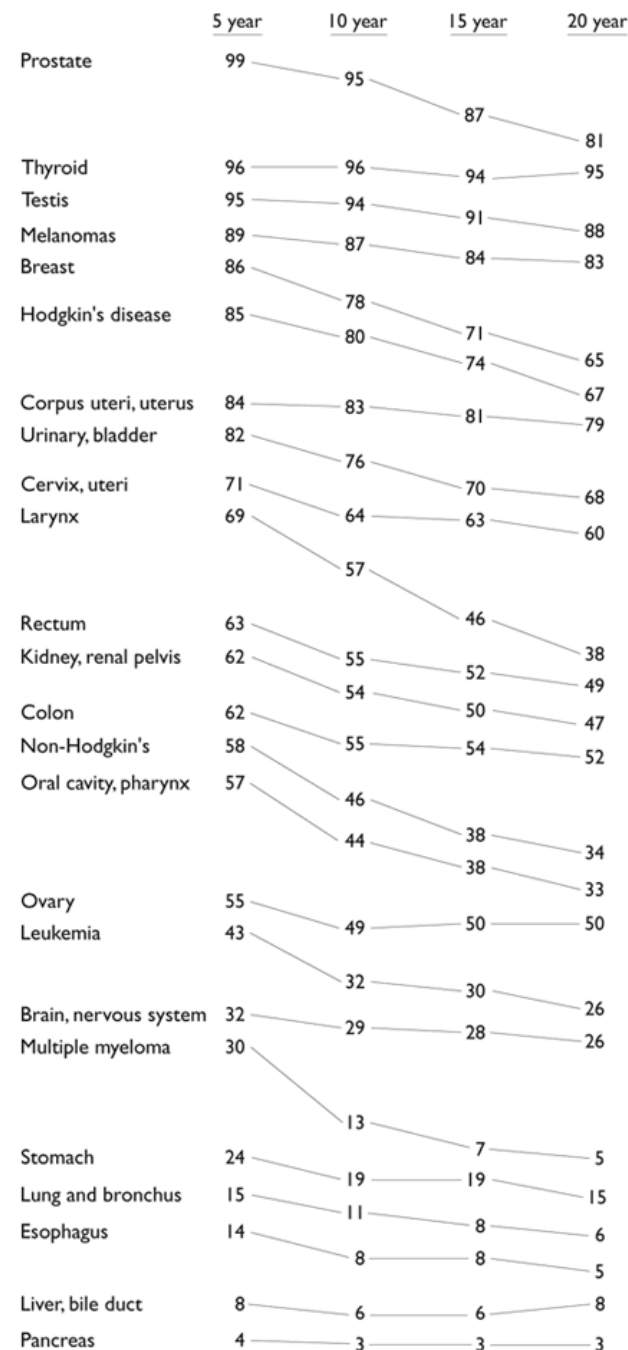
**Table 4: Most recent period estimates of relative survival rates, by cancer site**

**Estimates of relative survival rates, by cancer site**

|                       | % survival rates and their standard errors |          |          |          |
|-----------------------|--------------------------------------------|----------|----------|----------|
|                       | 5 year                                     | 10 year  | 15 year  | 20 year  |
| Prostate              | 98.8 0.4                                   | 95.2 0.9 | 87.1 1.7 | 81.1 3.0 |
| Thyroid               | 96.0 0.8                                   | 95.8 1.2 | 94.0 1.6 | 95.4 2.1 |
| Testis                | 94.7 1.1                                   | 94.0 1.3 | 91.1 1.8 | 88.2 2.3 |
| Melanomas             | 89.0 0.8                                   | 86.7 1.1 | 83.5 1.5 | 82.8 1.9 |
| Breast                | 86.4 0.4                                   | 78.3 0.6 | 71.3 0.7 | 65.0 1.0 |
| Hodgkin's disease     | 85.1 1.7                                   | 79.8 2.0 | 73.8 2.4 | 67.1 2.8 |
| Corpus uteri, uterus  | 84.3 1.0                                   | 83.2 1.3 | 80.8 1.7 | 79.2 2.0 |
| Urinary, bladder      | 82.1 1.0                                   | 76.2 1.4 | 70.3 1.9 | 67.9 2.4 |
| Cervix, uteri         | 70.5 1.6                                   | 64.1 1.8 | 62.8 2.1 | 60.0 2.4 |
| Larynx                | 68.8 2.1                                   | 56.7 2.5 | 45.8 2.8 | 37.8 3.1 |
| Rectum                | 62.6 1.2                                   | 55.2 1.4 | 51.8 1.8 | 49.2 2.3 |
| Kidney, renal pelvis  | 61.8 1.3                                   | 54.4 1.6 | 49.8 2.0 | 47.3 2.6 |
| Colon                 | 61.7 0.8                                   | 55.4 1.0 | 53.9 1.2 | 52.3 1.6 |
| Non-Hodgkin's         | 57.8 1.0                                   | 46.3 1.2 | 38.3 1.4 | 34.3 1.7 |
| Oral cavity, pharynx  | 56.7 1.3                                   | 44.2 1.4 | 37.5 1.6 | 33.0 1.8 |
| Ovary                 | 55.0 1.3                                   | 49.3 1.6 | 49.9 1.9 | 49.6 2.4 |
| Leukemia              | 42.5 1.2                                   | 32.4 1.3 | 29.7 1.5 | 26.2 1.7 |
| Brain, nervous system | 32.0 1.4                                   | 29.2 1.5 | 27.6 1.6 | 26.1 1.9 |
| Multiple myeloma      | 29.5 1.6                                   | 12.7 1.5 | 7.0 1.3  | 4.8 1.5  |
| Stomach               | 23.8 1.3                                   | 19.4 1.4 | 19.0 1.7 | 14.9 1.9 |
| Lung and bronchus     | 15.0 0.4                                   | 10.6 0.4 | 8.1 0.4  | 6.5 0.4  |
| Esophagus             | 14.2 1.4                                   | 7.9 1.3  | 7.7 1.6  | 5.4 2.0  |
| Liver, bile duct      | 7.5 1.1                                    | 5.8 1.2  | 6.3 1.5  | 7.6 2.0  |
| Pancreas              | 4.0 0.5                                    | 3.0 1.5  | 2.7 0.6  | 2.7 0.8  |

- Hermann Brenner, "Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis," The Lancet, 360 (October 12, 2002), 1131-1135.

# TABLES AND GRAPHS (2)



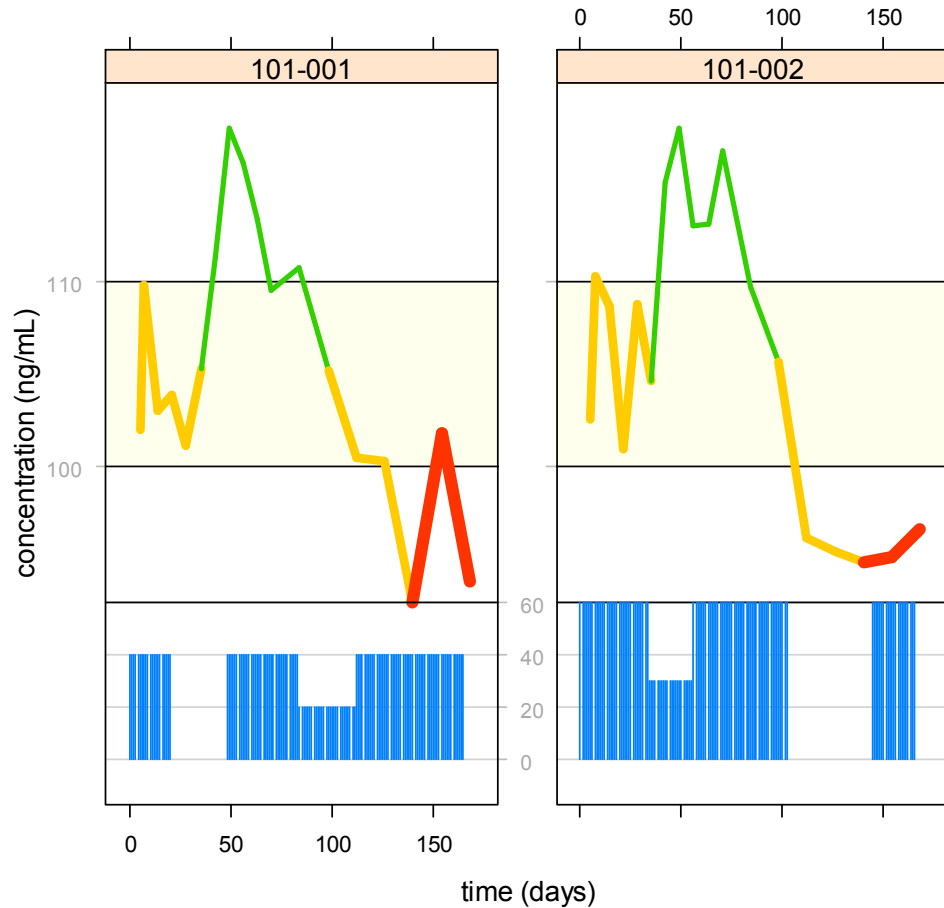
Cancer survival rates: tables, graphics, and PP

Edward Tufte, October 28, 2002

[http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg\\_id=0000Jr](http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0000Jr)

- Discussion of variations in: Preface, Krause and O'Connell (editors), A Picture is Worth a Thousand Tables

# A SIX-DIMENSIONAL SCATTERPLOT: PATIENT, BIOMARKER, RESPONSE, DOSE HISTORY, TIME, NORMAL RANGE



Subject (PID)

Biomarker concentration

Clinical Response:

Green: Response

Yellow: Stable disease

Red: Progressive disease

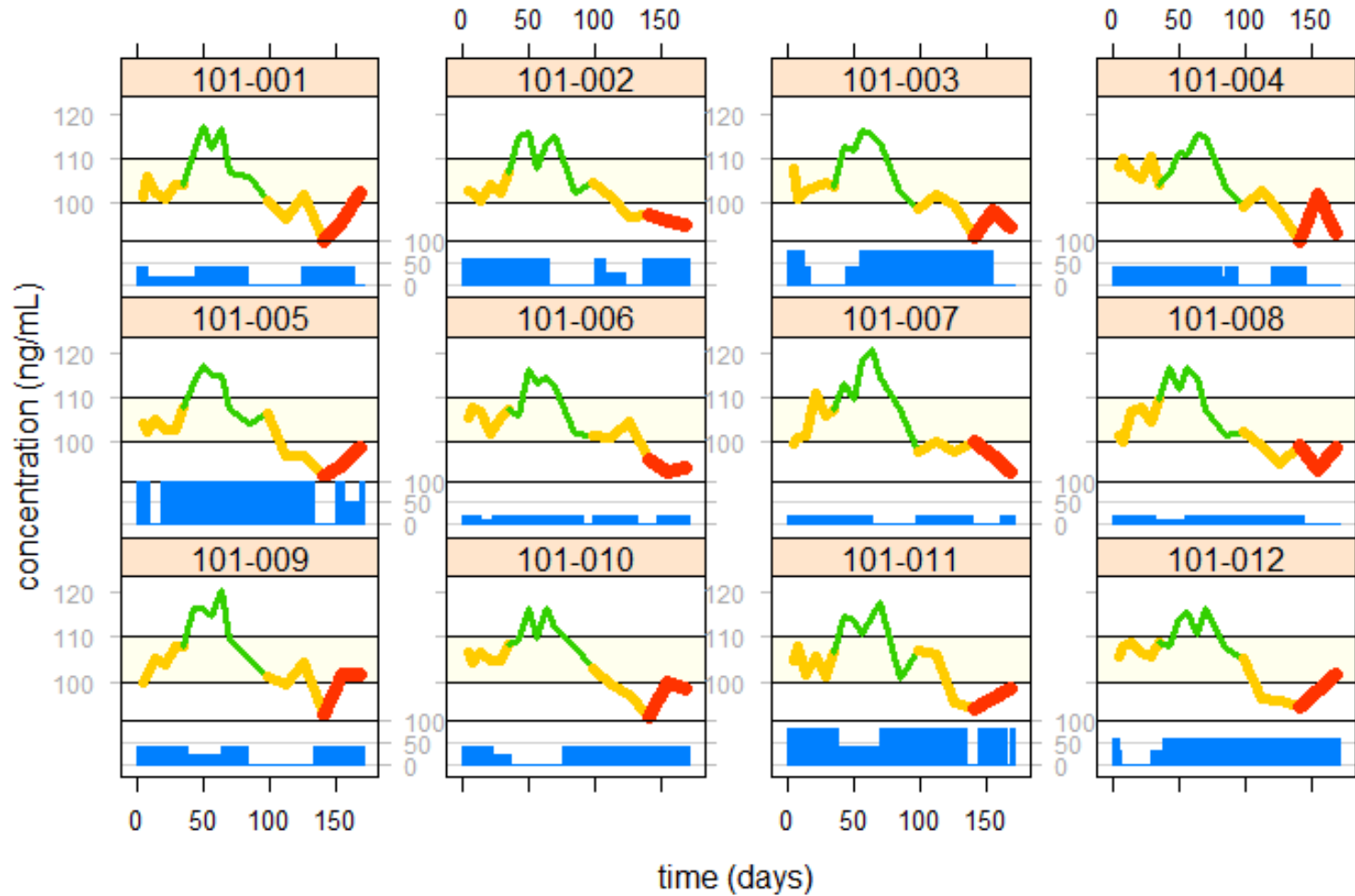
Thickness (monotone, thicker=worse, for b/w)

Normal range (shaded area)

Dose history (bars, axis)

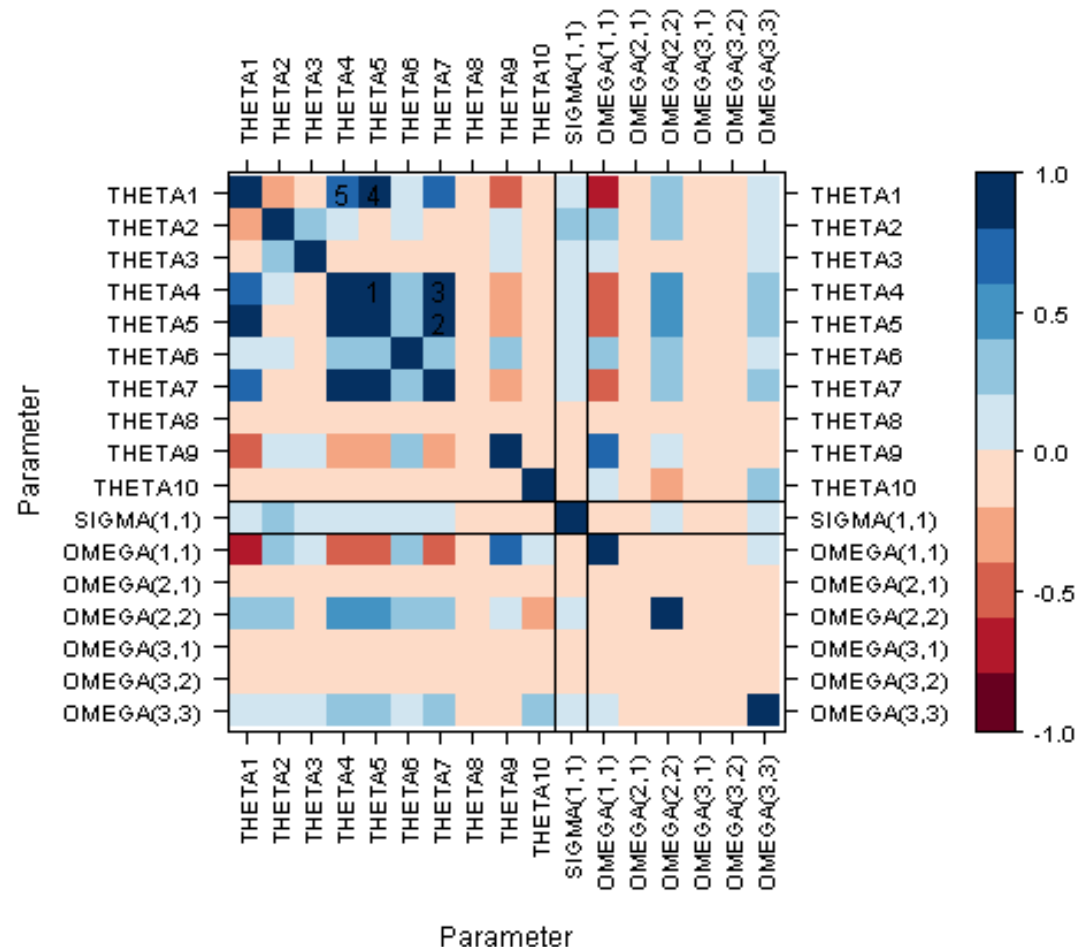
Time

# SIX-DIMENSIONAL SCATTERPLOT: 16 SUBJECTS



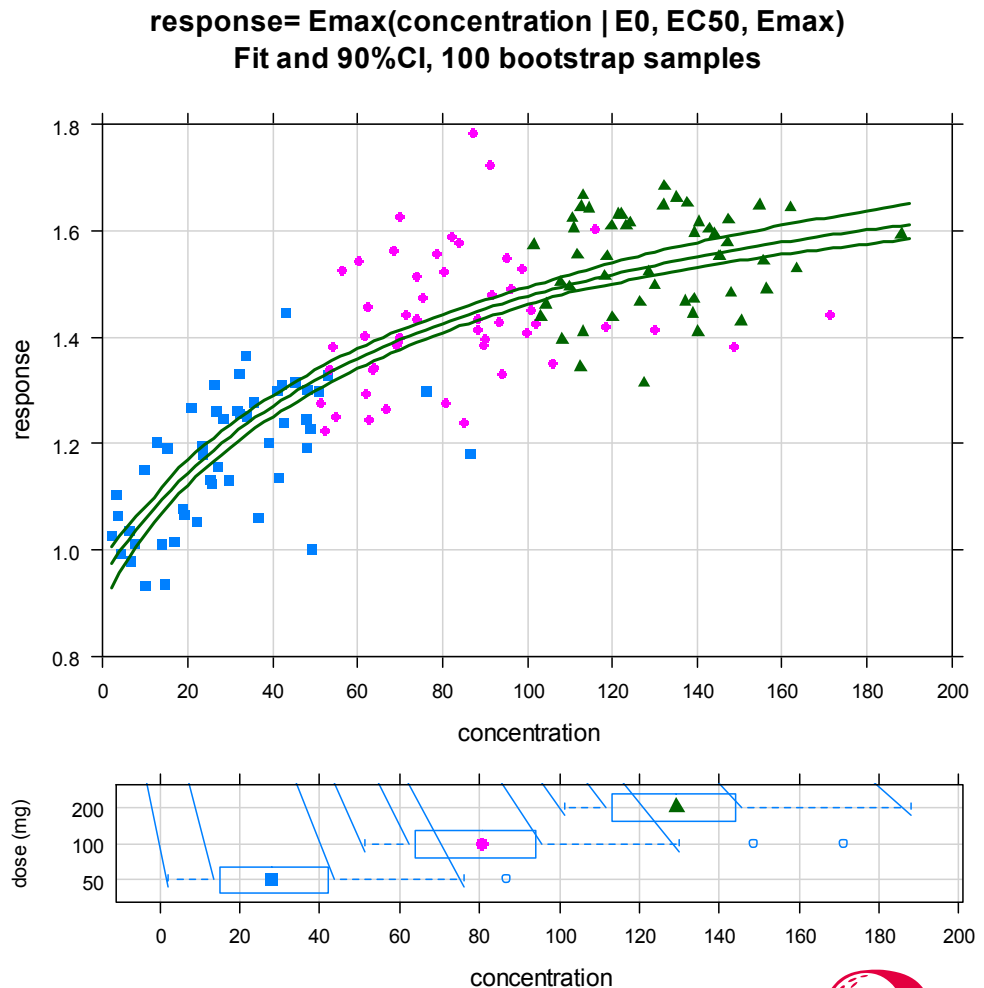
# PARAMETER CORRELATION

- ▶ Colors indicate correlation between a pair of parameters
- ▶ Blue: positive correlation
- ▶ Red: negative correlation
- ▶ Intense blue/ red: high positive/ negative correlation



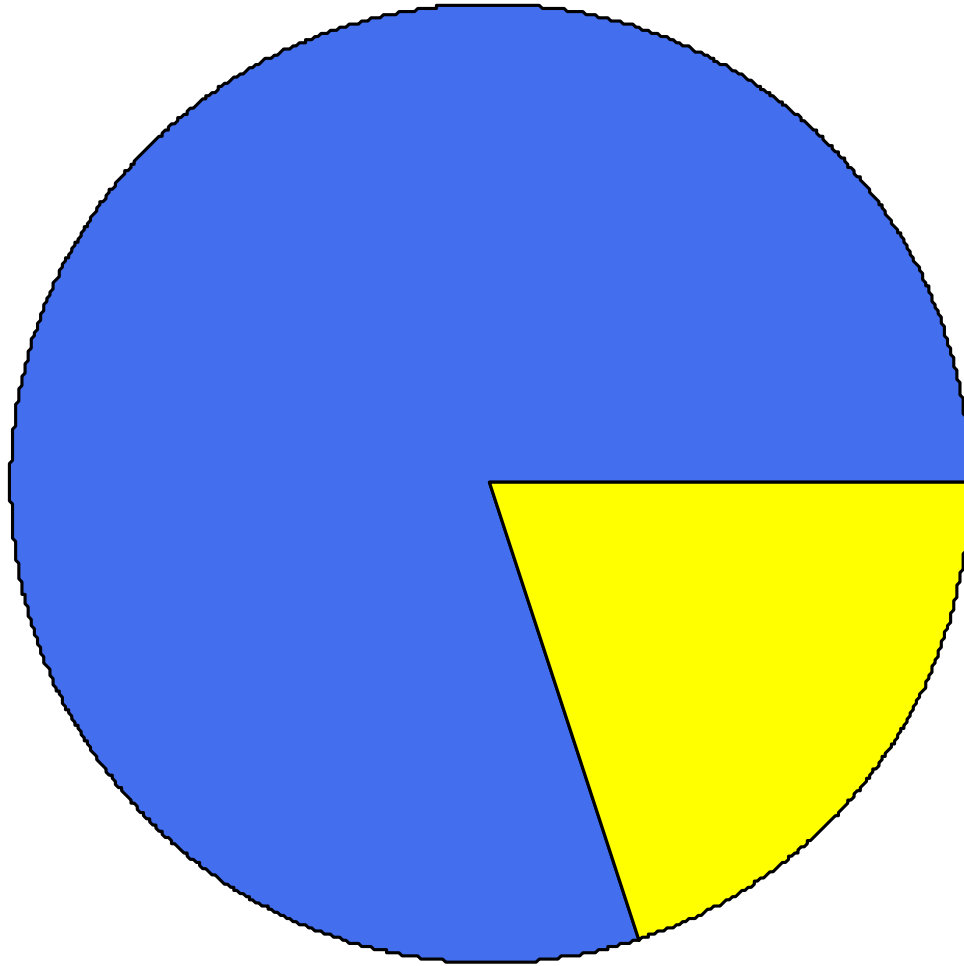
# CREATIVE IDEAS: CONCENTRATION-RESPONSE INTEGRATED WITH DOSE-RESPONSE

- ▶ Full information (top): concentration-response
  - ▶ Dose response (bottom): categorical variables
  - ▶ Aligned by superposition
  - ▶ Model fit and confidence interval
- 
- ▶ Note: there is no legend!
  - ▶ Reading off values is easier with a grid



# THE CHILDREN'S POOL

---



# SUMMARY

---

- ▶ Graphical presentation of PK, AE, labs, vital signs
  - **Can provide clear messages**
  - **Anchor the clinical study report**
  - **Rapidly and accurately** identify the information
  
- ▶ **Four key roles** for statistical graphics
  - Exploratory: Understand/explore the data, cleaning, outliers
  - Review: Medical review of population and patient level data, labs, AEs, vitals, concom meds, medical history
  - Submission: Clinical study reports and registration documents
  - Presentation: Scientific and marketing applications
  
- ▶ **Future (?)**
  - **Standardization** of data formats and graphics
    - Based on CDISC/SDTM
    - Across the industry

# READING ON VISUALIZATION

E. Tufte: **Envisioning Information**  
Beautiful evidence  
The Visual Display of Quantitative Information  
The cognitive style of PowerPoint

W.S. Cleveland: **Visualizing Data**

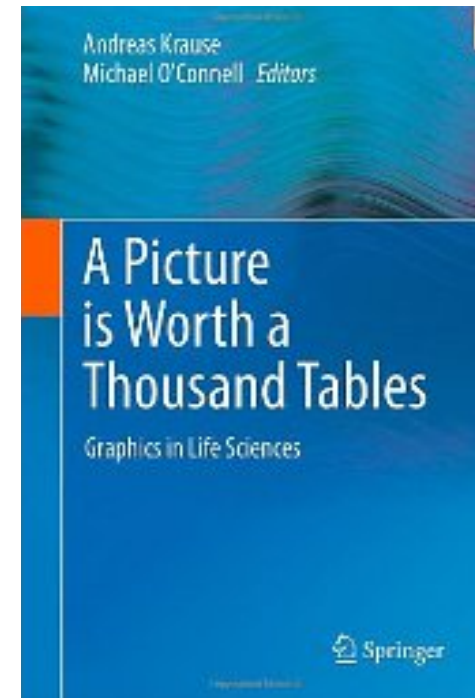
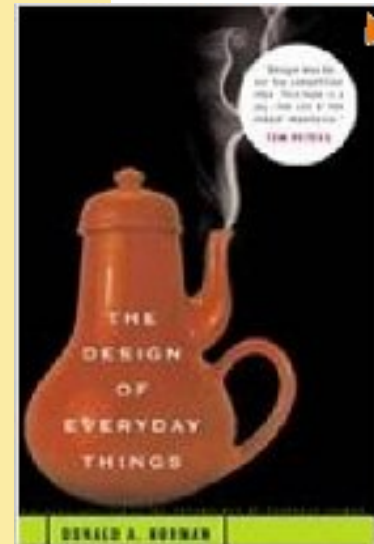
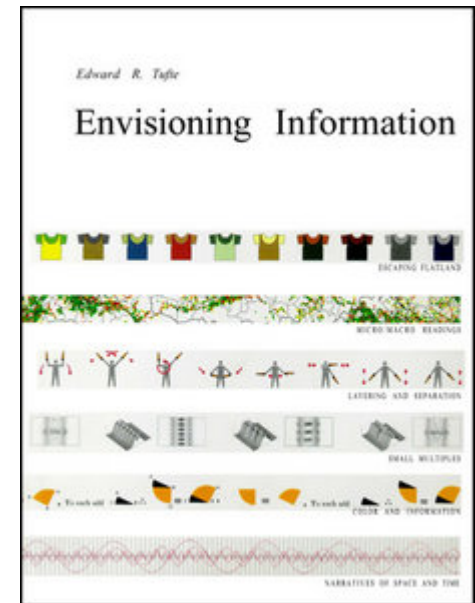
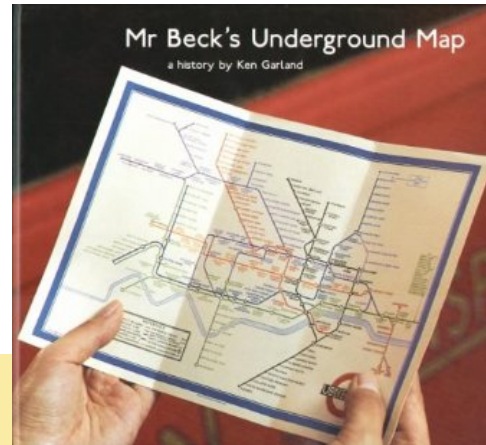
K. Garland: **Mr. Beck's Underground Map**

D.A. Norman: **The Design of Everyday's Things**

CTSpedia.org: Safety graphics and code

A. Krause, P. Lowe: **Visualization and Communication of PK/PD Models.**  
CPT PSP, May 2014

A. Krause, M. O'Connell (2012):  
**A Picture is Worth a Thousand Tables**  
**Phuse discount: 20% (until Aug 1, 2014):**  
**Use code KRAUSE20 on [www.springer.com](http://www.springer.com)**



# ACKNOWLEDGMENTS

---

- ▶ Michael O'Connell (Tibco)
- ▶ The authors of the book “A Picture Is Worth A Thousand Tables: Graphics in Life Sciences”
- ▶ Many colleagues for many good discussions
- ▶ Anonymous donors for providing examples (the good, the bad, and the ugly)

# THANK YOU.

# Today's Random Medical News

from the New England  
Journal of  
Panic-Inducing  
Gobbledygook

JIM BORGAN



# WHAT ARE GRAPHICS?

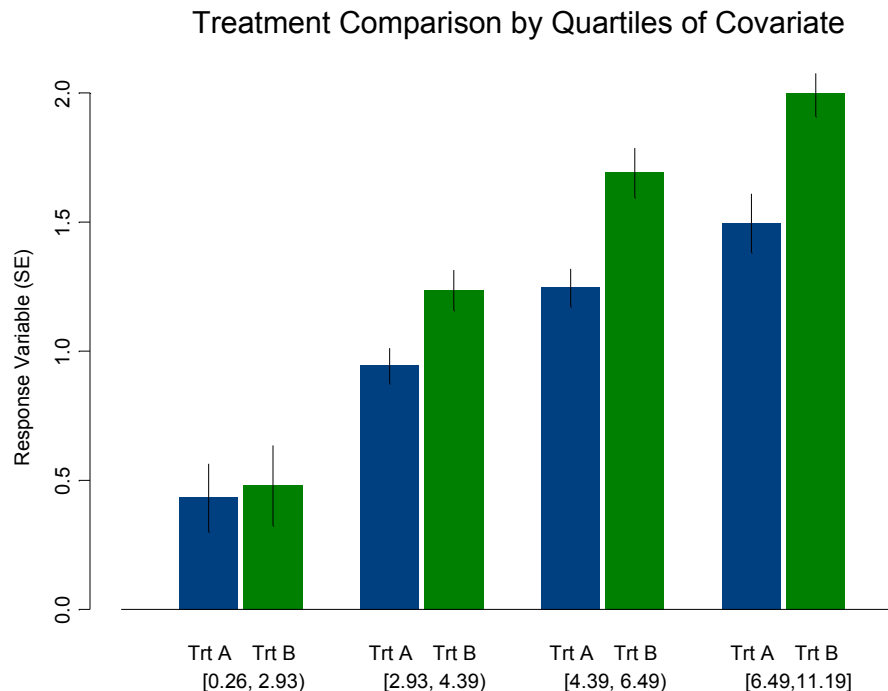
---

- ▶ Statistical graphics ... are information graphics in the field of statistics used to visualize quantitative data
- ▶ Good statistical graphics can provide a convincing means of communicating to others the underlying message that is present in the data
- ▶ **Graphical statistical methods have four objectives (Cleveland, 1994):**
  - The exploration of the content of a data set
  - The use to find structure in data
  - Checking assumptions in statistical models
  - Communicate the results of an analysis
- ▶ **Not using graphics increases the risk of missing a signal in the data**

Based on the Wikipedia entry for “statistical graphics”

# PRINCIPLE: CHOICE OF APPROPRIATE AGGREGATION LEVEL

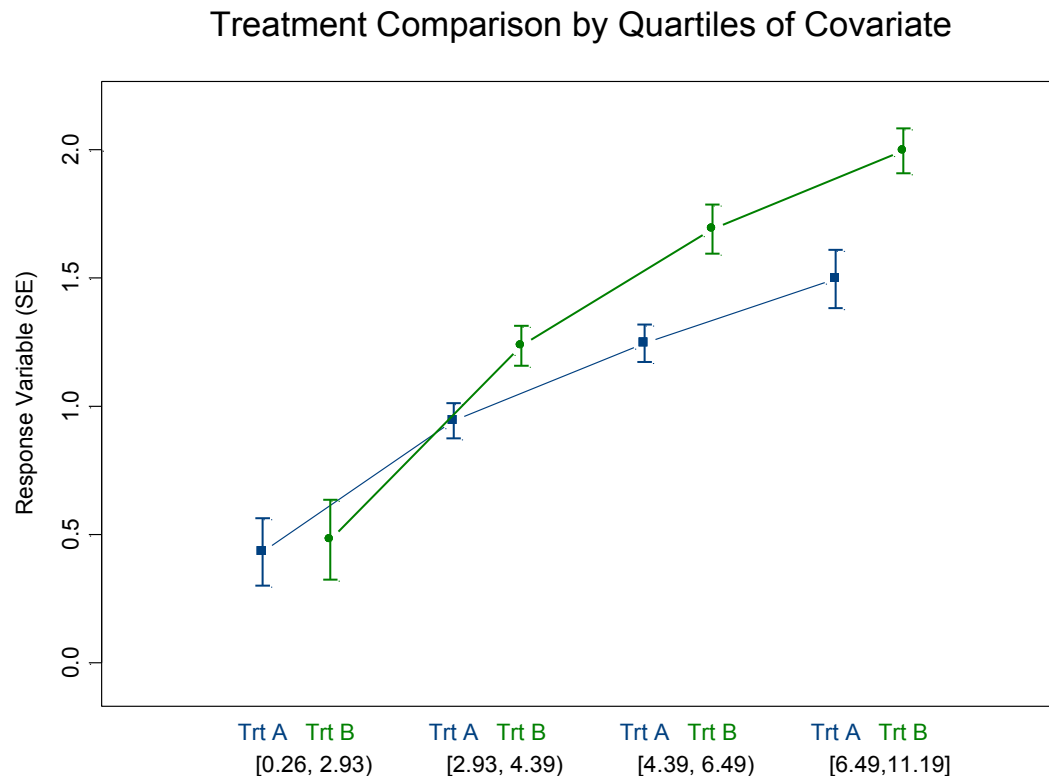
- ▶ Two treatments, A and B
- ▶ Observed range cut into four intervals depending on a covariate
- ▶ Poor data to ink ratio



The information of the bar plot is in the horizontal line on top of the bars.

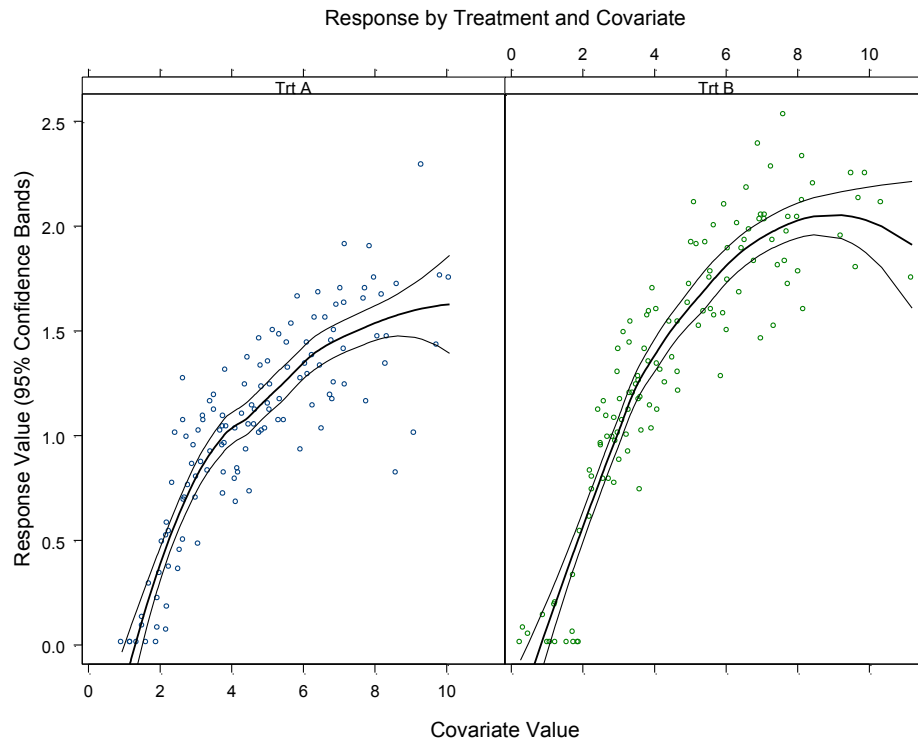
# PRINCIPLE: CHOICE OF APPROPRIATE AGGREGATION LEVEL

- ▶ Better data to ink ratio
- ▶ Allows comparison within and between groups (green and blue)
- ▶ Lines suggest connection and order



# PRINCIPLE: CHOICE OF APPROPRIATE AGGREGATION LEVEL

- ▶ Show all the data
- ▶ Added nonparametric smoother
- ▶ Allows judgment about patterns/ relationships, outliers, and more





# BACKUP SLIDES

---

## Contract of Employment (Individual Conditions)

*Between*

*And*

**Actelion Pharmaceuticals Ltd**

**Mr. Andreas Krause**

Dear Mr. Krause

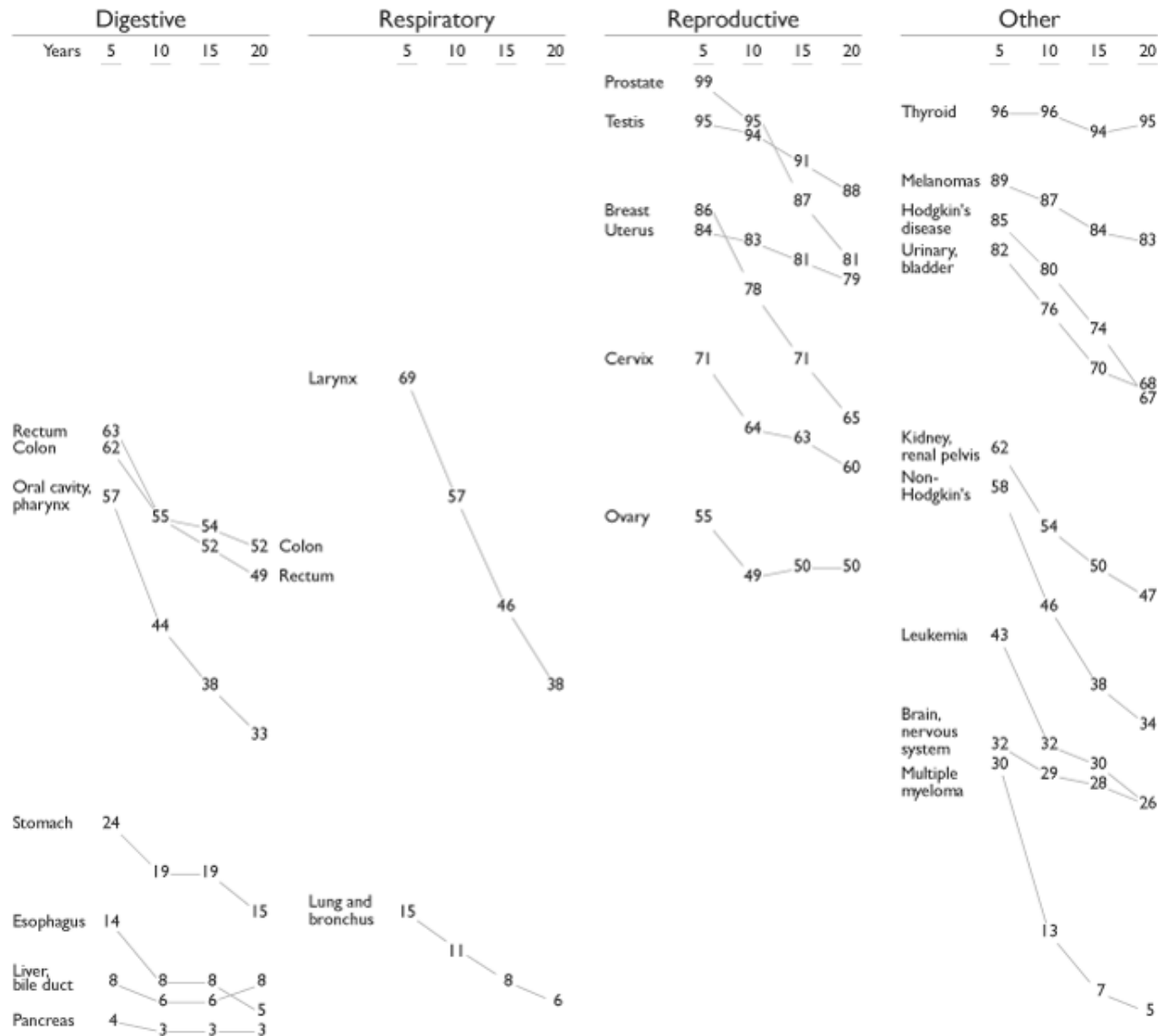
We are pleased to offer you the following individual terms of employment (subject to approval of your work permit by the authorities):

**Starting date:** 01.02.2008

**Place of Employment:** Gewerbestrasse 16, CH-4123 Allschwil

**Title:** Lead Scientist Modeling and Stimulation

# TABLES AND GRAPHS (3)



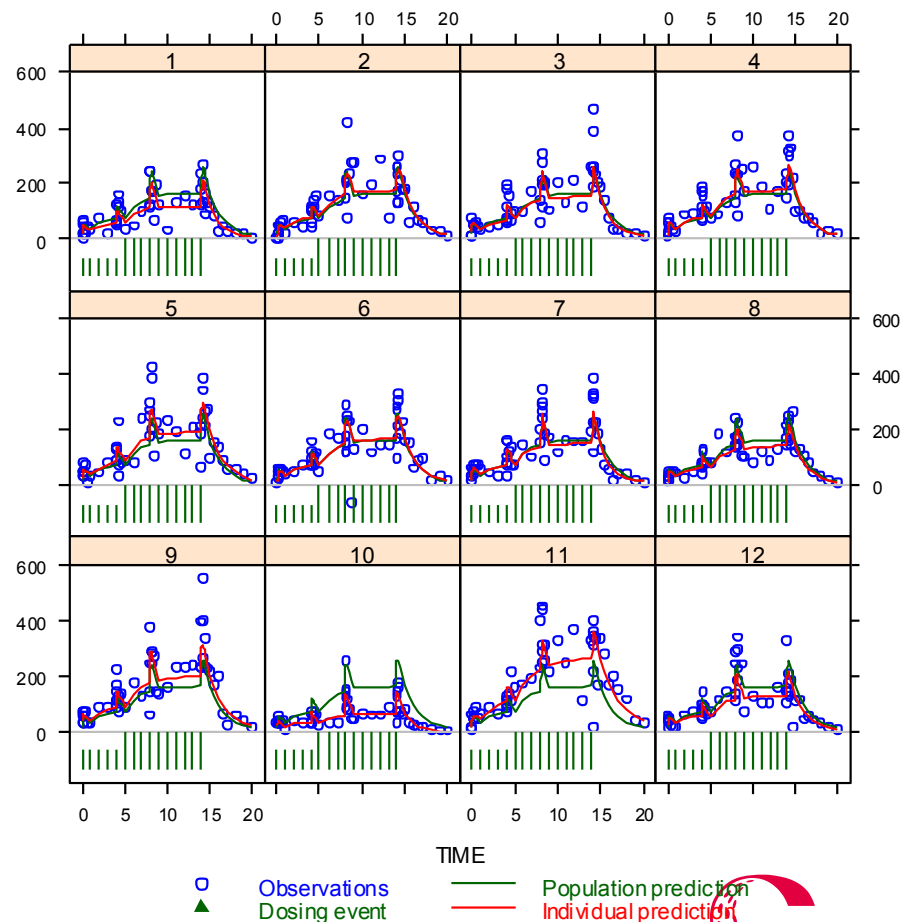
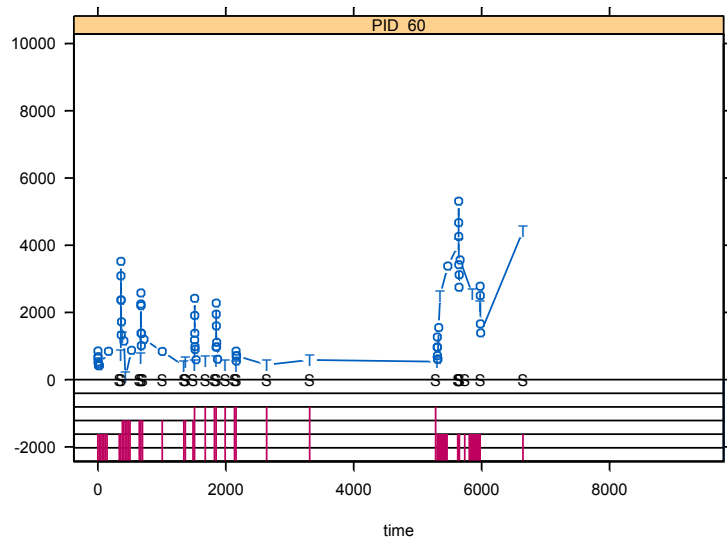
Dave Nash (Jan 19, 2006) in the discussion of Edward Tufte's chart.

[http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg\\_id=00004r](http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00004r)

recreation

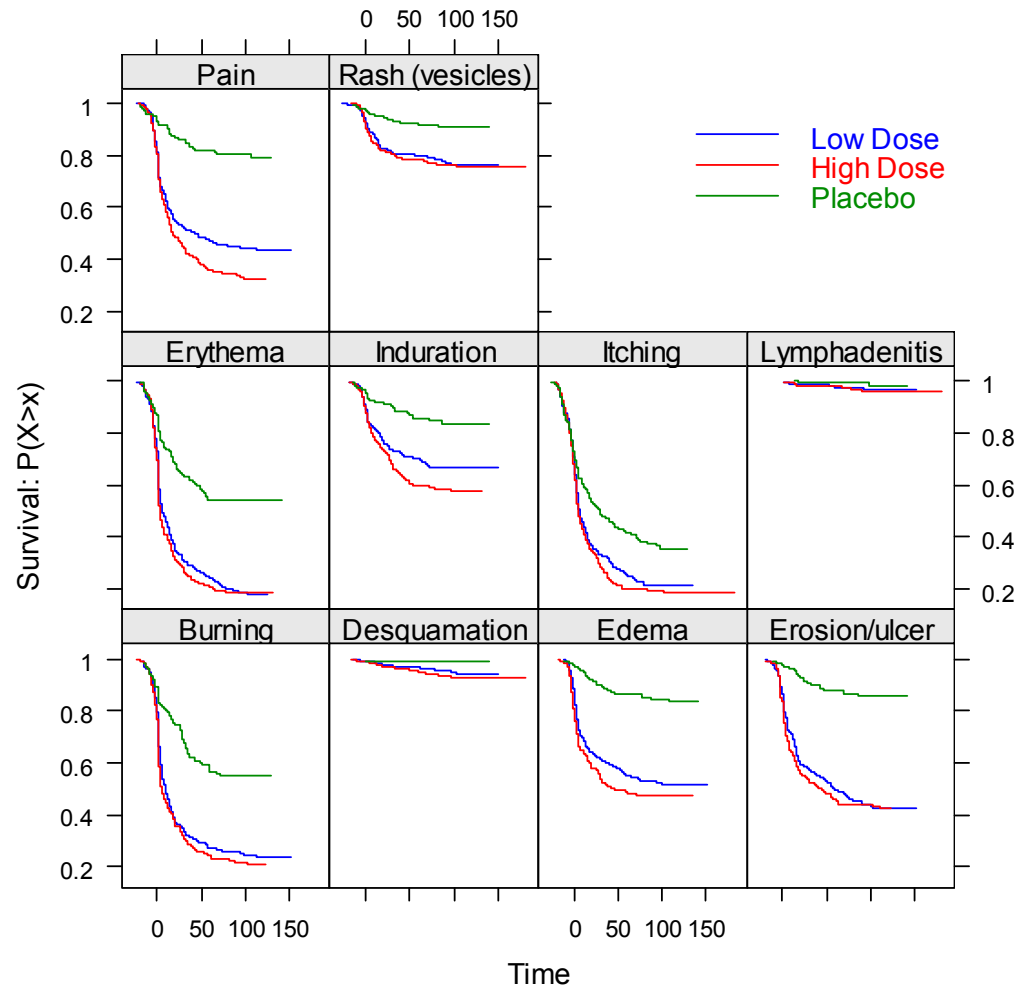
# CREATIVE IDEAS: CONCENTRATION-TIME PROFILES

- Concentration-Time Profile
- Trough levels labeled “T”
- Steady state labeled “S”



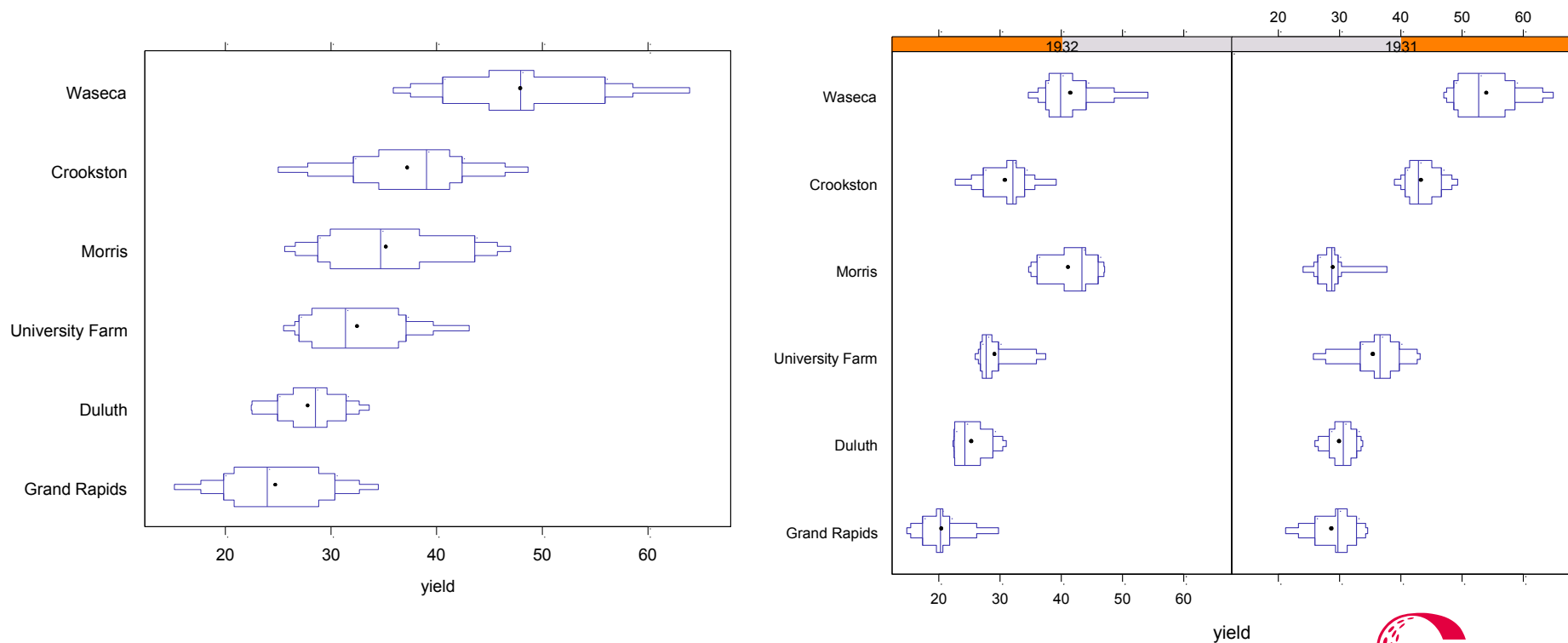
# CREATIVE IDEAS: KAPLAN-MEIER PLOT FOR SAFETY DATA

- ▶ Percentage of subjects without the adverse event (AE)
- ▶ Each panel shows one AE over time



# CREATIVE IDEAS: VIOLIN PLOTS

- Extension of the box-plot idea
- Each step denotes a quantile (10%, 20%, etc.)



# Today's Random Medical News

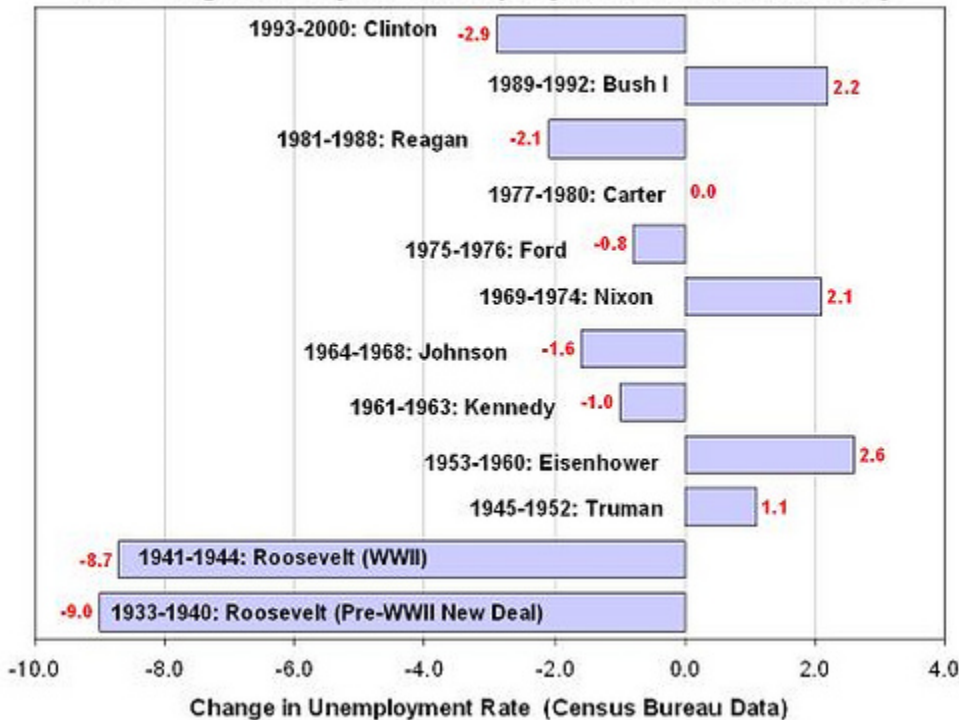
from the New England  
Journal of  
Panic-Inducing  
Gobbledygook

JIM BORGAN



# PERCENTAGES

FDR's Pre-WWII New Deal:  
The Largest Drop In Unemployment Rate in History



- ▶ ROOSEVELT PRE-WWII NEW DEAL
  - 1932 Unemployment Rate: 23.6%
  - 1940 Unemployment Rate: 14.6%
  - Rate Change: -9.0

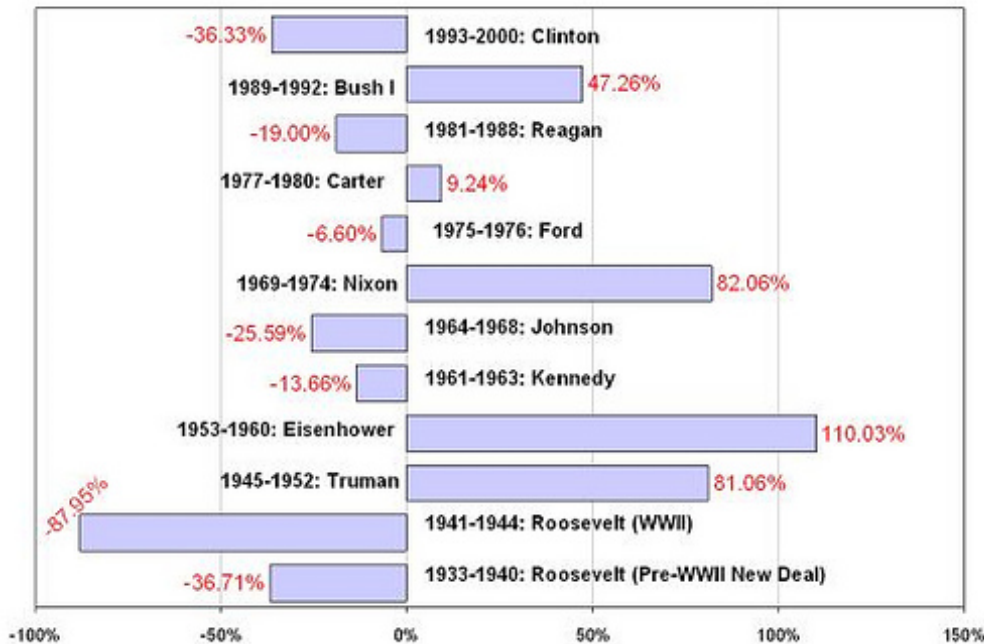
- ▶ Clinton years
  - 1993: 6.9%
  - 2000: 4.0%

- ▶ One cannot reduce the unemployment rate by 14.6 percent if it currently is at 6.9 percent!

- ▶ Do not even think about percentages of a percentage.

# PERCENTAGES

Percentage Change in Unemployment By President



- ▶ Do not even think about percentages of a percentage!
- ▶ But guess what: here it is!
- ▶ 6.9% to 4.0% unemployment is a drop by 36%!
- ▶ Presentation bias
  - Suggesting one-sided view

[http://www.huffingtonpost.com/david-sirota/the-forgotten-math-pre-ww\\_b\\_155728.html](http://www.huffingtonpost.com/david-sirota/the-forgotten-math-pre-ww_b_155728.html)

Aug 22, 2009

# BEWARE OF PERCENTAGES

---

- ▶ Percentages are always tricky!
- ▶ Consider this:
  - “Our drug” shows a survival rate of 80%
  - The competitor drug shows a survival rate of 75%
  - So “our drug” has a survival rate that is 5% higher
  - Or is the competitor  $(80-75)/80$  6% worse?
  - Or is our compound  $(80-75)/75 = 7\%$  better?
  - Switch to “risk” or death rate:
    - “our drug”: 20%, the competitor 25%
    - The death rate of the competitor is thus 25% higher!
- ▶ That line of argument is actually used (guess which!)

# PERCENTAGES

---

- Analogy of the previous in a clinical context:
- (That kind of advertisement is actually not unusual)
- Patients treated with Femara for five years following surgery had a **13% reduced risk of death** compared with those treated with tamoxifen (p=0.08).
- Five-year survival rates were
  - **87.9% for women treated with Femara only;**
  - **86.2% for those treated with two years of tamoxifen followed by three years of Femara**
- Difference in **death rate**:  $(87.9\% - 86.2\%) / (100\% - 86.2\%) = 13\% !!$

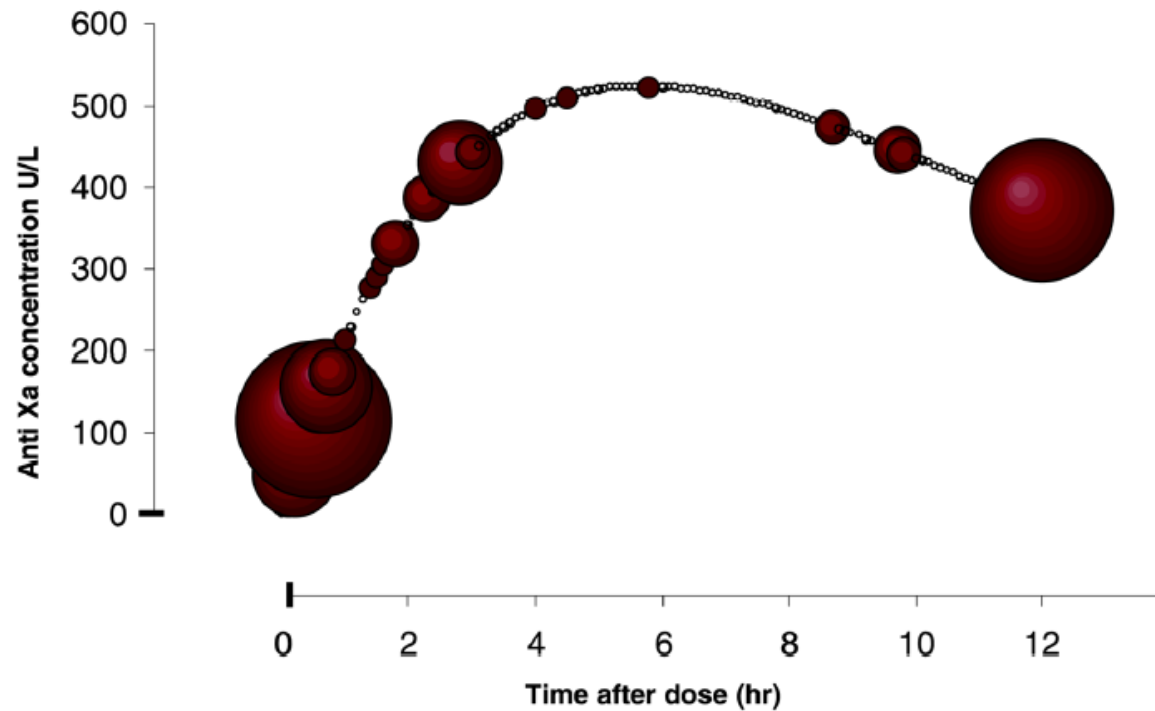
► [http://professional.cancerconsultants.com/oncology\\_main\\_news.aspx?id=42967](http://professional.cancerconsultants.com/oncology_main_news.aspx?id=42967), Aug 22, 2009

# PERCENTAGES

---

- ▶ Analogy of the previous in a clinical context:
- ▶ (That kind of advertisement is actually not unusual)
- ▶ Patients treated with Femara for five years following surgery had a **13% reduced risk of death** compared with those treated with tamoxifen (p=0.08).
- ▶ Five-year **survival rates** were 87.9% for women treated with Femara only; 86.2% for those treated with two years of tamoxifen followed by three years of Femara
- ▶ Difference in **death rate**:  $(87.9\% - 86.2\%) / (100\% - 86.2\%) = 13\% !!$

[http://professional.cancerconsultants.com/oncology\\_main\\_news.aspx?id=42967](http://professional.cancerconsultants.com/oncology_main_news.aspx?id=42967), Aug 22, 2009



- ▶ Graphs show the number of samples taken from patients over time, lined up on a typical concentration-time curve.
- ▶ Time axis: time intervals that correspond to the “blobs” have to be guessed
- ▶ “blobs” overlay
- ▶ Visual perception depends on the order of drawing of the blobs
- ▶ Data (number of samples) is proportional to what: diameter, area, other?
- ▶ If anything, they should be proportional to the area, since this is how the human eye interprets it.

- ▶ Going to a 3-dimensional effect does not improve the clarity of what is shown

# FINAL WORD

---

“There are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don’t know. But there are also unknown unknowns. There are things we do not know we don’t know.”



Donald H. Rumsfeld  
Defense Department Briefing  
Feb 12, 2002

# A GLIMPSE OF THE FUTURE?

---

- ▶ Statistical Graphics plug in to clinical data seamlessly
  - Standardized clinical data sources (e.g., CDISC)
- ▶ Statistical Graphics are suggested based on data types
  - Using the clinical graphics taxonomy
- ▶ A Statistical Graphics Palette is standardized across industry and authorities
  - Simple interactions between industry and authorities
  - Safety and the Critical Path Initiative are making this reality

# THANK YOU.