Data Sets for Mixed-Effects Models

Help Files
Version 3.0
June 1999

by Douglas M. Bates and José C. Pinheiro

University of Wisconsin — Madison and Bell Labs, Lucent Technologies

This document describes the data sets included in the NLMEDATA subdirectory of the nlme distribution. We have adopted naming conventions for the columns in the data frames and for the levels in the factors within the data frames, especially the grouping factor. These naming conventions are not required. We find that they help us remember the roles of the different columns or variables. They also can provide more meaningful labels on the plots.

The title of each section in this document gives the name of the corresponding groupedData object in the NLMEDATA subdirectory, followed by a short description of the data. The formula stored with the data is given next followed by a more detailed description of each of the columns. The first column generates the response, the second column is the primary covariate and the third column is the primary grouping factor. Other covariates and grouping factors are then listed. Usually we use lower case for the names of the columns except for the primary grouping factor which we capitalize.

1 Assay — Bioassay on cell culture plate

These data, courtesy of Rich Wolfe and David Lansky, Searle Inc. come from a bioassay run on a 96 well cell culture plate. The assay is performed using a split-block design. The 8 rows on the plate are labeled A-H from top to bottom, while the 12 columns on the plate are labeled 1-12 from left to right. Only the central 60 wells of the plate are used for the bioassay (the intersection of rows B-G and columns 2-11). There are two blocks in the design: Block 1 contains columns 2-6 and Block 2 contains columns 7-11. Within each block, 6 samples are randomly assigned to rows and 5 (serial) dilutions are randomly assigned to columns. The response variable is the logarithm of the optical density. The cells are treated with a compound that they metabolize to produce the stain. Only live cells can make the stain, so the optical density is a measure of the number of cells that are alive and healthy.

Columns The display formula for these data is

```
logDens ~ 1 | Block
```

based on the columns named:

`logDens`: log-optical density.

`Block`: a factor identifying the block where the wells are measured.

`sample`: a factor identifying the sample corresponding to the well, varying from “a” to “f”.

`dilut`: a factor indicating the dilution applied to the well, varying from 1 to 5.

2 BodyWeight — Body weight growth in rats

Hand and Crowder (1996) describe data on the body weights of rats measured over 64 days. These data also appear in Table 2.4 of Crowder and Hand (1990). The body weights of the rats (in grams) are measured on day 1 and every seven days thereafter until day 64, with an extra measurement on day 44. There are three groups of rats, each on a different diet.

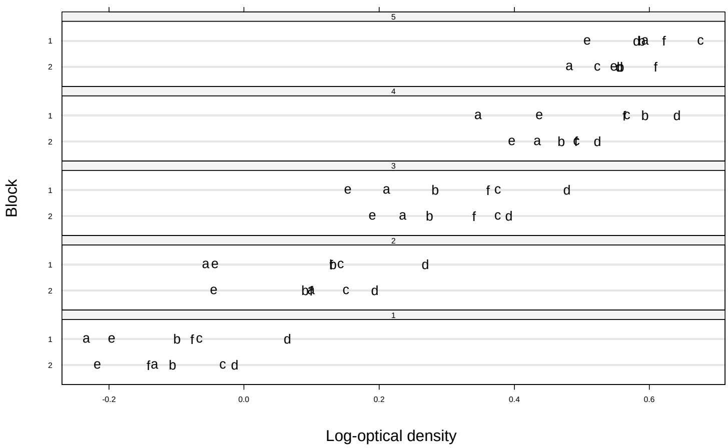

Figure 1: *Log-optical density measured on the central 60 wells of a cell culture plate. The wells are divided into two blocks with 6 rows and 5 columns, with samples being assigned to rows and dilutions to columns. Panels in the plot correspond to the serial dilutions and symbols refer to the samples.*

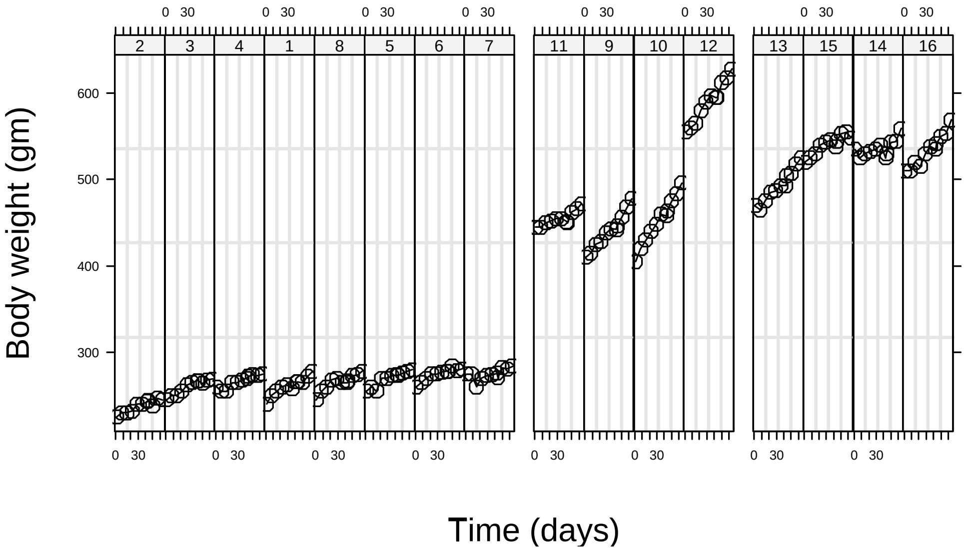

Figure 2: *Body weights of rats measured over a period of 64 days. The rats are divided into three groups on different diets.*

Columns The display formula for these data is

```
weight ~ Time | Rat
```

based on the columns named:

`weight`: body weight of the rat (grams).

`time`: time at which the measurement is made (days).

`Rat`: a factor identifying the rat whose weight is measured.

`Diet`: a factor indicating the diet the rat receives.

3 ChickWeight — Early growth of baby chicks

Hand and Crowder (1996) describe data on the body weights of chicks over time. These data also appear in Example 5.3 of Crowder and Hand (1990). The body weights of the chicks are measured at birth and every second day thereafter until day 20. They are also measured on day 21. There are four groups on chicks on different protein diets.

Columns The display formula for these data is

```
weight ~ time | Chick
```

based on the columns named:

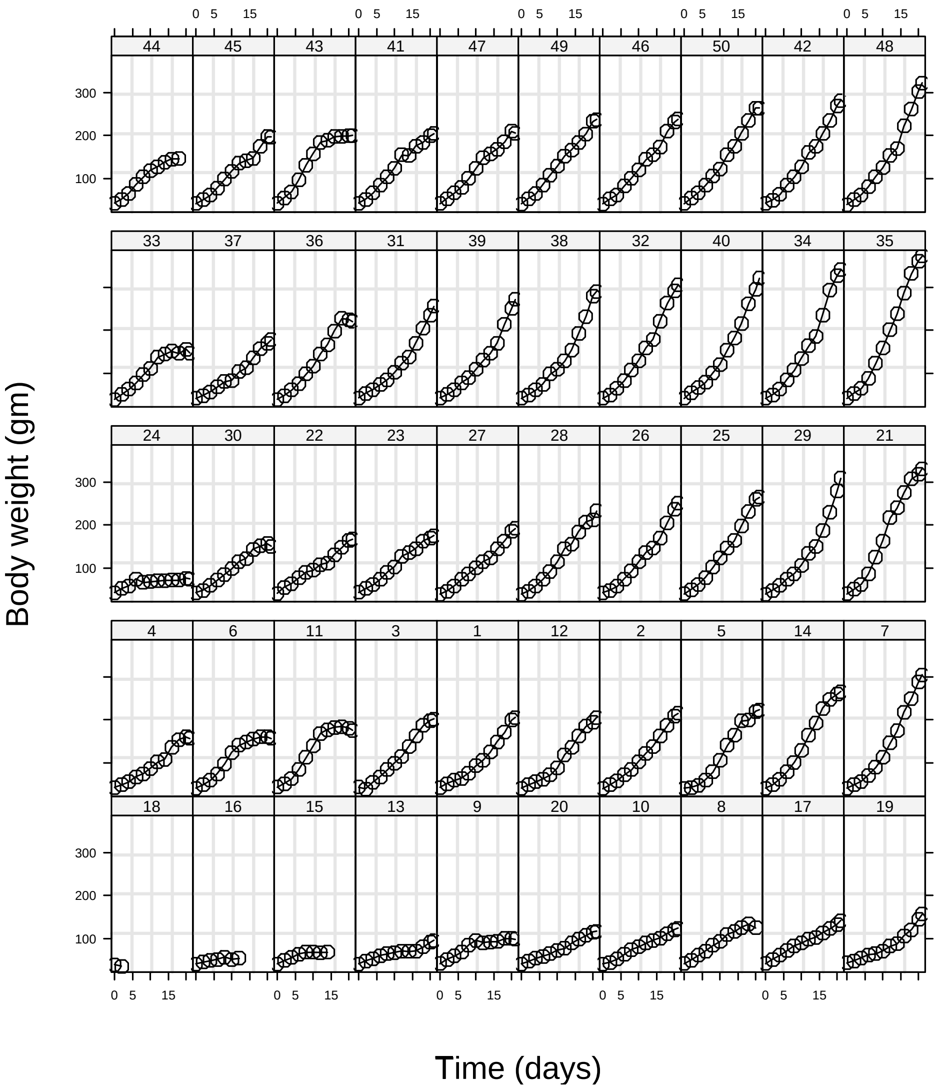

Figure 3: *Body weights of baby chicks measured at birth and every second day thereafter until day 20. The chicks are divided into four groups on different protein diets.*

`weight`: body weight of the chick (gm).

`time`: time at which the measurement is made (days).

`Chick`: a factor identifying the chick whose weight is measured.

`Diet`: a factor indicating the protein diet the chick is receiving.

Models Hand and Crowder (1996) suggest using a linear or quadratic model for weight gain over time. Figure 3 could also suggest a four-parameter logistic model.

4 CO2 — Carbon Dioxide uptake

Potvin, Lechowicz and Tardif (1990) describe an experiment on the cold tolerance of a C_4 grass species, *Echinochloa crus-galli*. The CO_2 uptake of six plants from Quebec and six plants from Mississippi was measured at several levels of ambient CO_2 concentration. Half the plants of each type were chilled overnight before the experiment was conducted. The data are shown in Figures 4 and 5.

Columns The display formula for these data is

```
uptake ~ conc | Plant
```

based on the columns named:

`uptake`: carbon dioxide uptake rate (μmol/m^2 sec).

`conc`: ambient concentration of carbon dioxide (mL/L).

`Plant`: a factor giving a unique identifier for each plant.

`Type`: origin of the plant - Quebec or Mississippi.

`Treatment`: treatment - chilled or nonchilled.

and the default formula for display is

Models Potvin et al. (1990) suggest using a modified form of the asymptotic regression model.

5 Dialyzer — High Flux Hemodialyzer

Vonesh and Carter (1992) describe data measured on high flux hemodialyzers to assess their *in vivo* ultrafiltration characteristics. The ultrafiltration rates (in ml/hr) of 20 high flux dialyzers were measured at 7 different transmembrane pressures (in dmHg). The *in vitro* evaluation of the dialyzers used bovine blood at flow rates of either 200 dl/min or 300 dl/min. The data, shown in Figure 6, are also analyzed in Littell, Milliken, Stroup and Wolfinger (1996, §8.2).

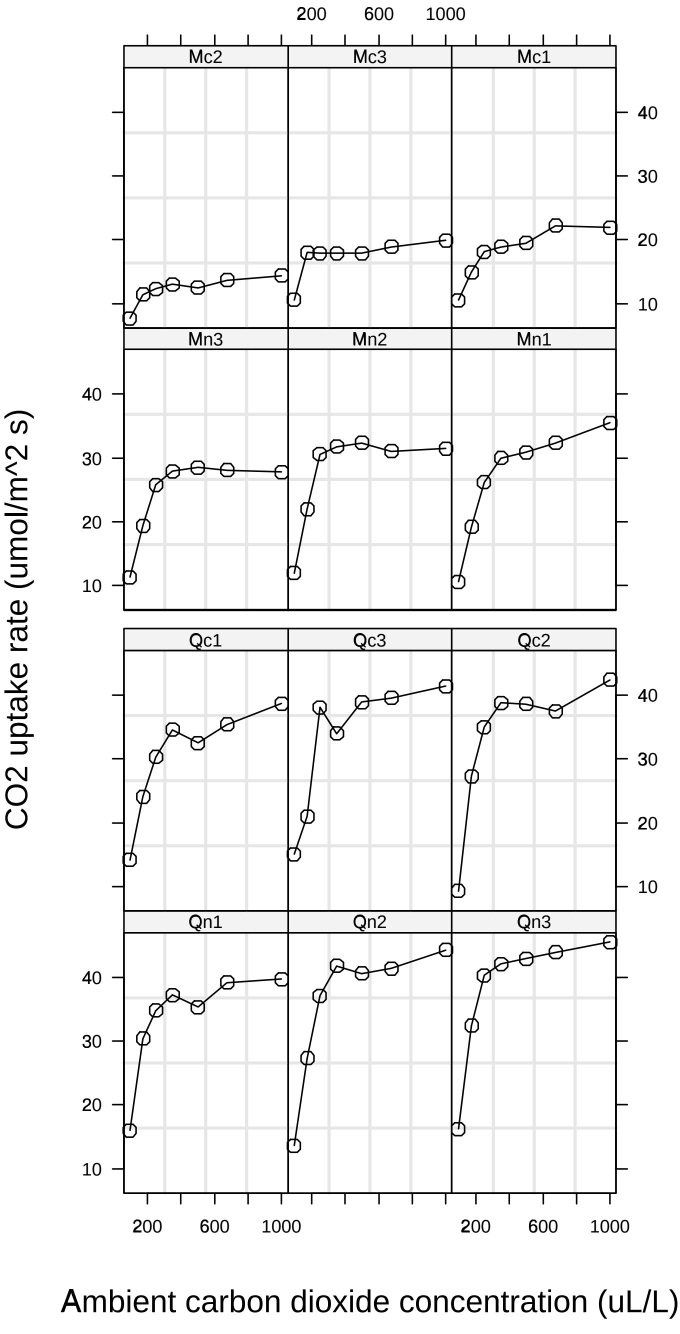

Figure 4: *CO_2 uptake versus ambient CO_2 for Echinochloa crus-galli plants, six from Quebec and six from Mississippi. Half the plants of each type were chilled overnight before the measurements were taken.*

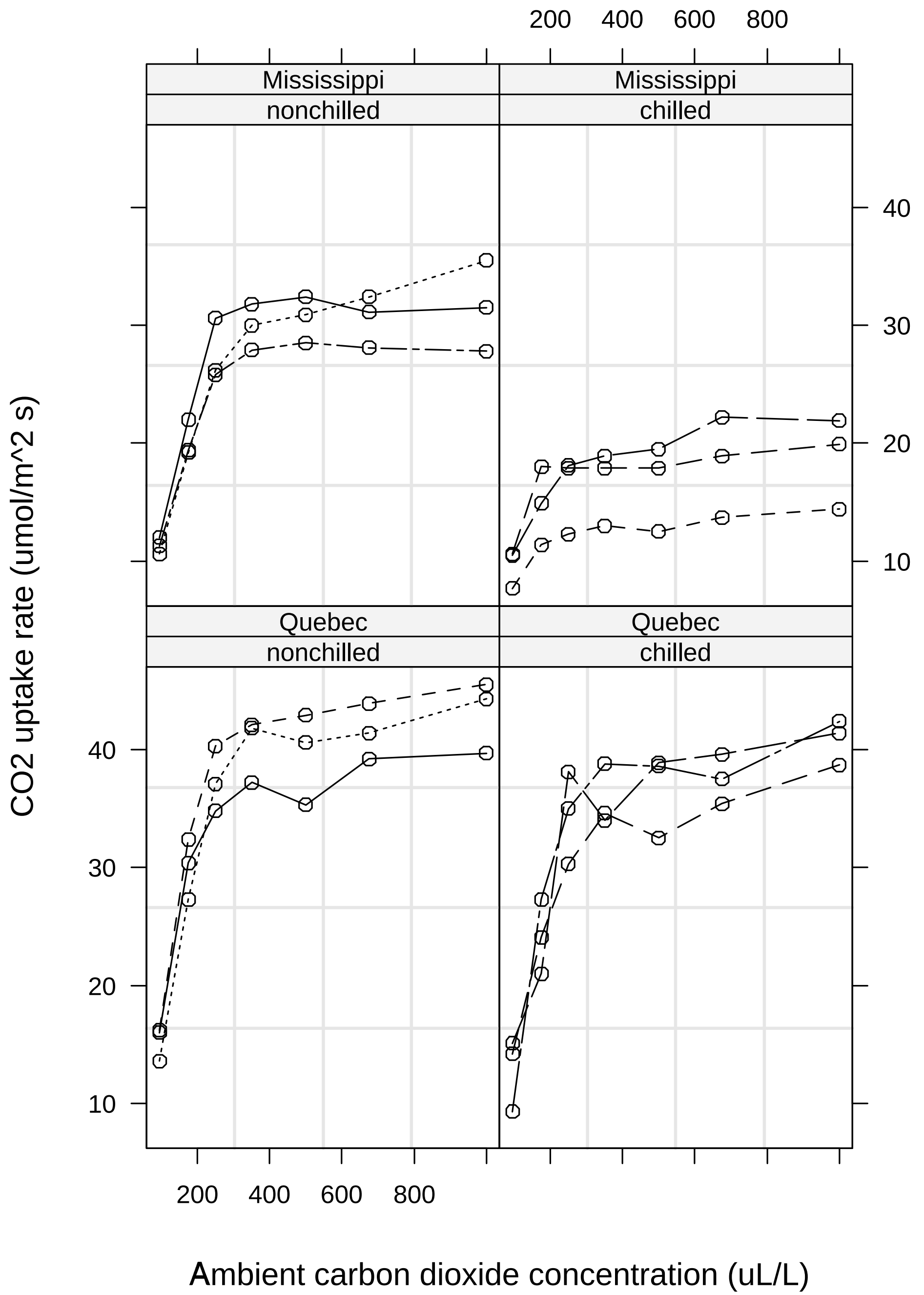

Figure 5: *CO_2 uptake versus ambient CO_2 by treatment and type for Echinochloa crus-galli plants, six from Quebec and six from Mississippi. Half the plants of each type were chilled overnight before the measurements were taken.*

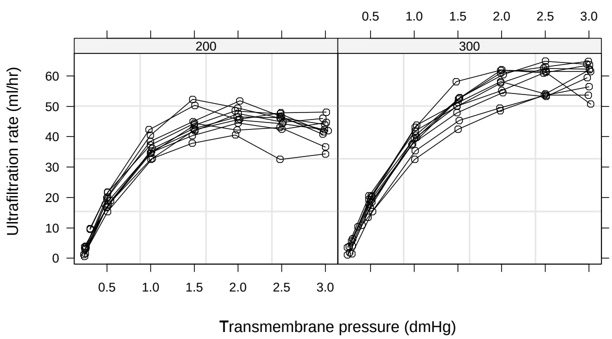

Figure 6: *Hemodialyzer ultrafiltration rates (in ml/hr) measured at 7 different transmembrane pressures (in dmHg) on 20 high flux dialyzers.* In vitro *evaluation of dialyzers based on bovine blood flow rates of 200 dl/min and 300 dl/min.*

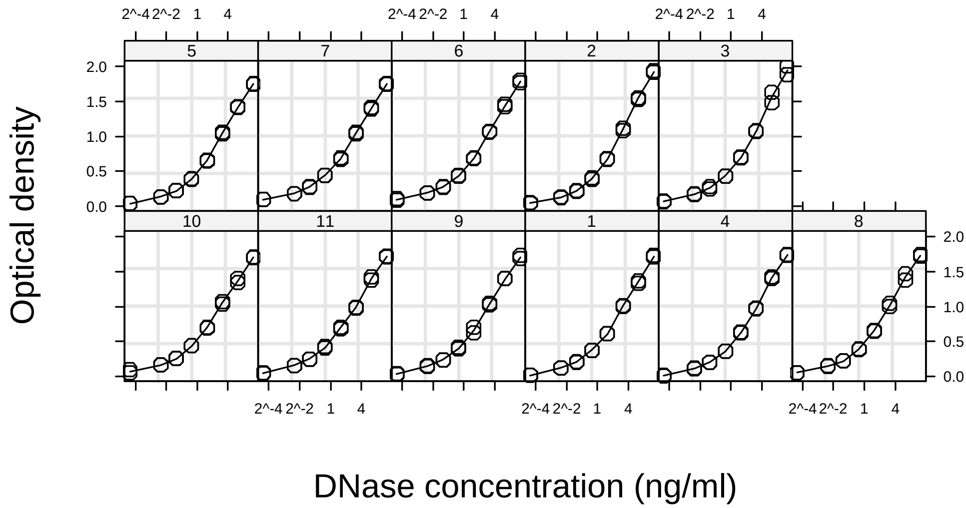

Figure 7: *Optical density versus DNase concentration for 11 runs of an assay. The concentration is shown on a logarithmic scale. For each run there are duplicate measurements at each concentration.*

Columns The display formula for these data is

```
rate ~ pressure | Subject
```

based on the columns named:

`rate`**:** hemodialyzer ultrafiltration rate (ml/hr).

`pressure`**:** transmembrane pressure (dmHg).

`Subject`**:** a factor giving a unique identifier for each subject.

`QB`**:** bovine blood flow rate (dl/min) - 200 or 300.

`index`**:** index of observation within subject - 1 through 7.

6 DNase — Assay Data for the Protein DNase

Davidian and Giltinan (1995, § 5.2.4, p. 134) describe data, shown in Figure 7, obtained during the development of an ELISA assay for the recombinant protein DNase in rat serum.

Columns The display formula for these data is

```
density ~ conc | Run
```

based on the columns named:

`density`: the measured optical density in the assay. Duplicate optical density measurements were obtained.

`conc`: the known concentration of the protein.

`Run`: a factor giving the run from which the data were obtained.

Models Davidian and Giltinan (1995) use the four-parameter logistic model with these data modelling the optical density as a logistic function of the logarithm of the concentration.

7 Earthquake — Earthquake intensity

These data, shown in Figure 8, are measurements recorded at available seismometer locations for 23 large earthquakes in western North America between 1940 and 1980. They were originally given in Joyner and Boore (1981); are mentioned in Brillinger (1987); and are analyzed in section 11.4 of Davidian and Giltinan (1995).

Columns The display formula for these data is

```
accel ~ distance | Quake
```

based on the columns named:

`accel`: maximum horizontal acceleration observed (g).

`distance`: the distance from the seismological measuring station to the epicenter of the earthquake (km).

`Quake`: a factor indicating the earthquake on which the measurements were made.

`Richter`: the intensity of the earthquake on the Richter scale.

`soil`: soil condition at the measuring station - either soil or rock.

8 Fatigue — Cracks caused by metal fatigue

These data are given in Lu and Meeker (1993) where they state

> We obtained the data in Table 1 visually from figure 4.5.2 on page 242 of Bogdanoff and Kozin (1985).

The data represent the growth of cracks in metal for 21 test units. An initial notch of length 0.90 inches was made on each unit which then was subjected to several thousand test cycles. After every 10,000 test cycles the crack length was measured. Testing was stopped if the crack length exceeded 1.60 inches, defined as a failure, or at 120,000 cycles. The data are shown in Figure 9

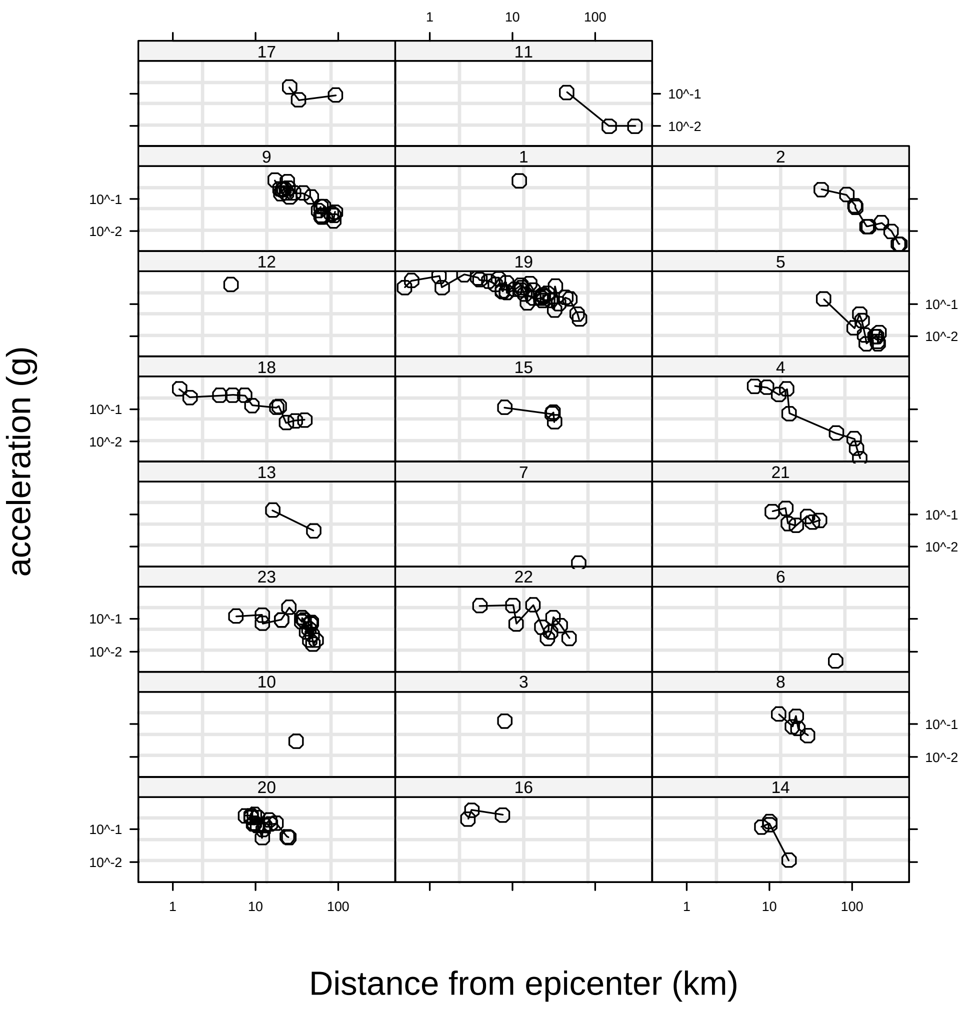

Figure 8: *Lateral acceleration versus distance from the epicenter for 23 large earthquakes in western North America. Both the acceleration and the distance are on a logarithmic scale. Earthquakes of greatest intensity as measured on the Richter scale are in the uppermost panels.*

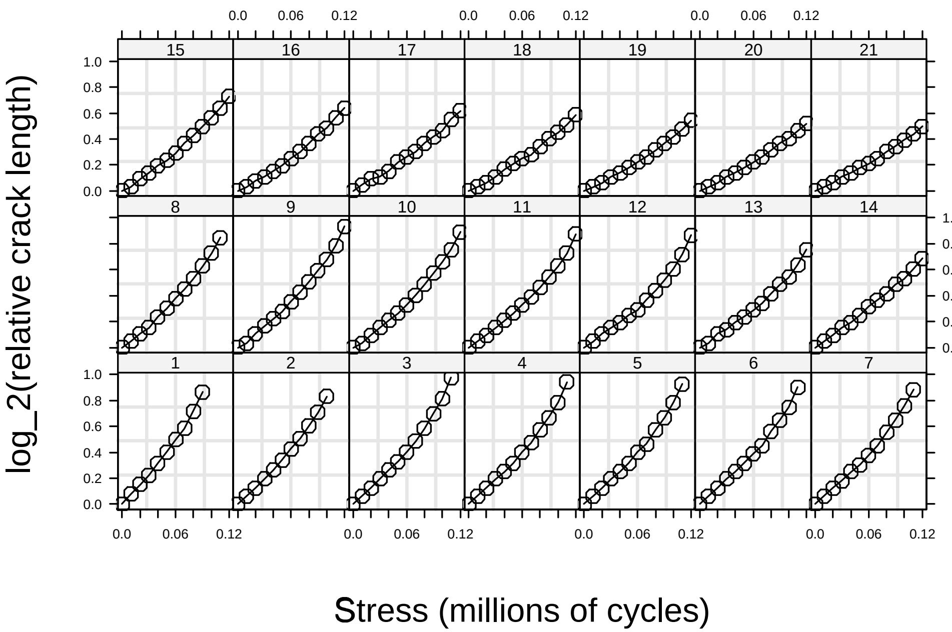

Figure 9: *Crack length, relative to initial notch length, versus number of cycles for the test unit. In all test units the crack was grown from an initial notch of 0.90 inches.*

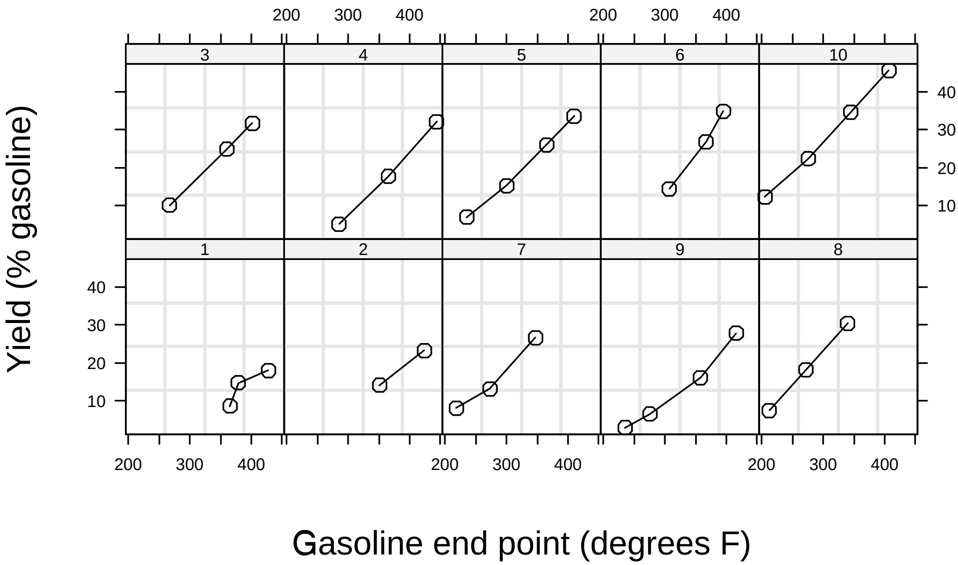

Figure 10: *Gasoline yield versus gasoline end point for ten samples of crude oil*

Columns The display formula for these data is

```
relLength ~ cycles | Path
```

based on the columns named:

`length`: relative crack length (dimensionless).

`cycles`: number of test cycles at which the measurement is made (millions of cycles).

`Path`: an ordered factor giving the test path (or test unit) number. The order is in terms of increasing failure time or decreasing terminal crack length.

9 Gasoline — Refinery yield of gasoline

Prater (1955) provides data on crude oil properties and gasoline yields. These data were used in Hader and Grandage (1958) to illustrate multiple regression analysis. Atkinson (1985) also uses these data to illustrate the use of diagnostics in multiple regression analysis. Three of the covariates — API, vapor, and ASTM — measure characteristics of the crude oil used to produce the gasoline. The other covariate — endpoint — is a characteristic of the refining process. Daniel and Wood (1980) notice that the covariates characterizing the crude oil occur in only ten distinct groups and conclude that the data represent responses measured on ten different crude oil samples.

In Figure 10 we present these data using these groupings. The display formula for these data is

```
yield ~ endpoint | Sample
```

based on the columns named:

`yield`: percentage of crude oil converted to gasoline after distillation and fractionation

`endpoint`: temperature (°F) at which all the gasoline is vaporized

`Sample`: inferred crude oil sample number

`API`: crude oil gravity (°API)

`vapor`: vapor pressure of the crude oil (lbf/in^2)

`ASTM`: crude oil 10% point ASTM — the temperature at which 10% of the crude oil has become vapor.

Models Based on the relationship shown in Figure 10 we begin with a simple linear relationship between yield and endpoint

10 Gun – Methods for firing naval guns

Hicks (1993, p. 180) reports data from an experiment on methods for firing naval guns. Gunners of three different physiques (slight, average, and heavy) tested two firing methods. Both methods were tested twice by each of nine teams of three gunners with identical physique. The response was the number of rounds fired per minute. The data, displayed in Figure 11, are used in Chambers and Hastie (1992, §5.2, p. 152) to illustrate the methods available in S-PLUSfor analyzing ANOVA models.

Columns The display formula for these data is

```
rounds ~ 1 | Team
```

based on the columns named:

`rounds`: the number of rounds fired per minute.

`Team`: a factor giving the three-gunner team.

`Method`: firing method - M1 or M2.

`Physique`: gunners' physique - Slight, Average, or Heavy.

Models

```
Gun.lme <- lme(rounds ~ Method + Physique, random = ~ 1, data = Gun)
```

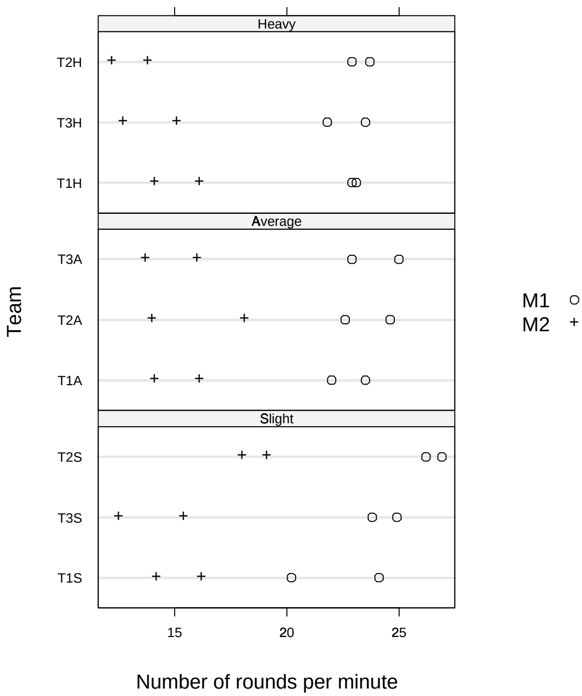

Figure 11: *Number of rounds fired per minute by three-gunner teams with one of three possible physiques, using two different firing methods.*

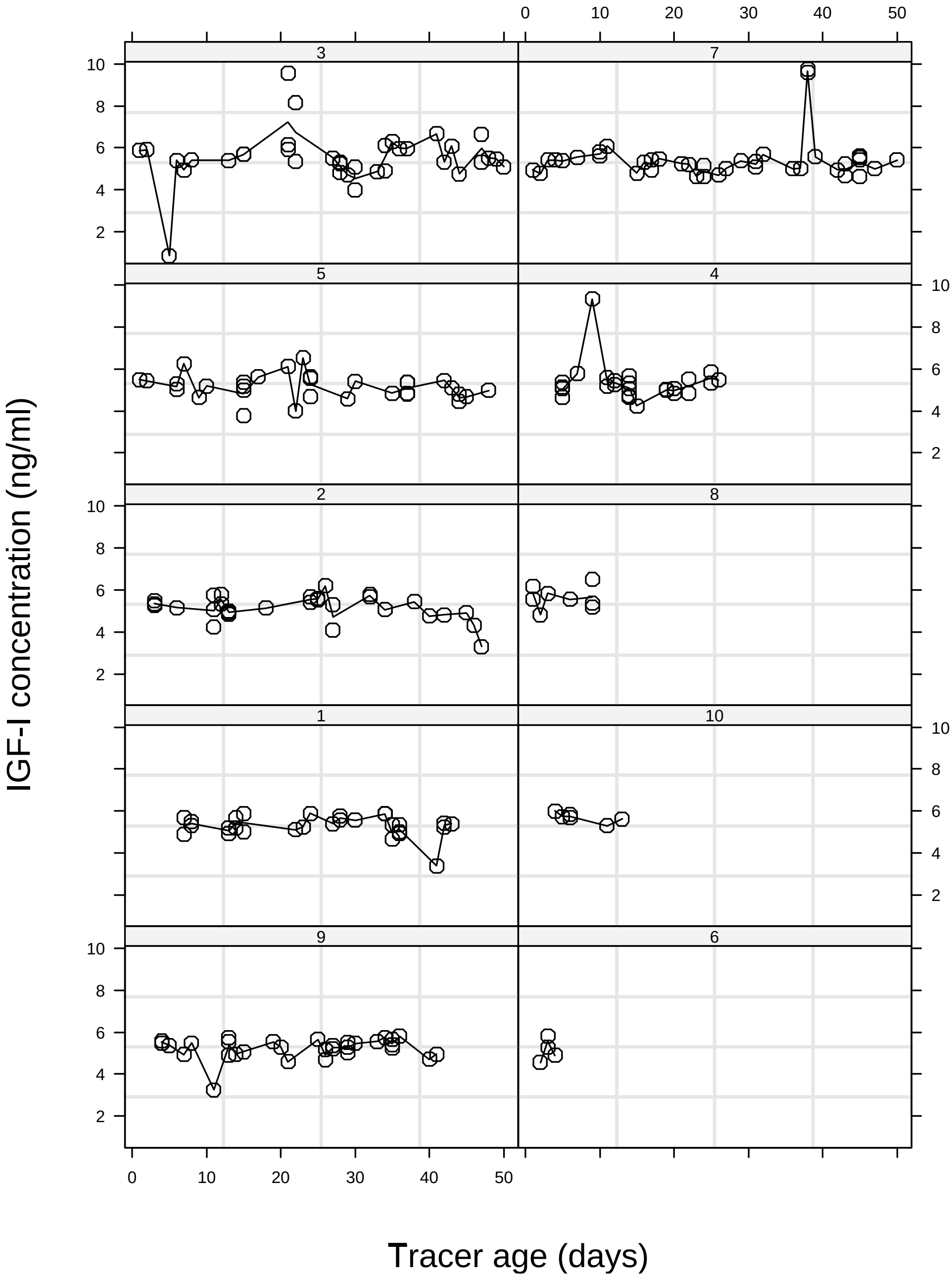

Figure 12: *Estimated concentration of the protein Insulin-like Growth Factor (IGF-I) versus age of radioactive tracer for 10 lots of tracer.*

11 IGF – Radioimmunoassay of IGF-I protein

Davidian and Giltinan (1995, § 3.2.1, p. 65) describe data, shown in Figure 12, obtained during quality control radioimmunoassays for ten different lots of radioactive tracer used to calibrate the Insulin-like Growth Factor (IGF-I) protein concentration measurements.

Columns The display formula for these data is

```
conc ~ age | Lot
```

based on the columns named:

`conc`: the estimated concentration of IGF-I protein, in ng/ml.

`age`: the age (in days) of the radioactive tracer.

`Lot`: a factor giving the radioactive tracer lot.

Models

```
IGF.lis <- lmList(conc ~ age, data = IGF)
IGF.lme <- lme(IGF.lis)
```

12 Indometh — Indomethicin kinetics

Kwan, Breault, Umbenhauer, McMahon and Duggan (1976) present data on the plasma concentrations of Indomethicin following intravenous injection. There are six different subjects in the experiment. The sampling times, ranging from 15 minutes post injection to 8 hours post injection, are the same for each subject. The data, presented in Figure 13, are analyzed in Davidian and Giltinan (1995, § 2.1)

The display formula for these data is

```
conc ~ time | Subject
```

based on the columns named:

`conc`: observed plasma concentration of indomethicin (mcg/ml).

`time`: time at which the sample was drawn (hr. post injection).

`Subject`: a factor indicating the subject from whom the sample is drawn.

13 Loblolly — Growth of Loblolly pine trees

Kung (1986) presents data, shown in Figure 14, on the growth of Loblolly pine trees.

The display formula for these data is

```
height ~ age | Seed
```

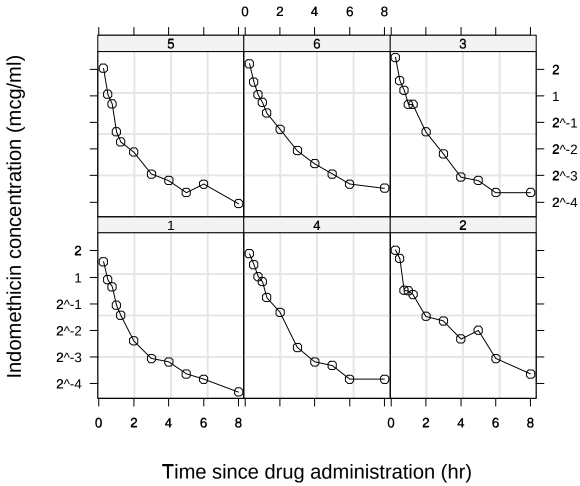

Figure 13: *Concentration of indomethicin over time*

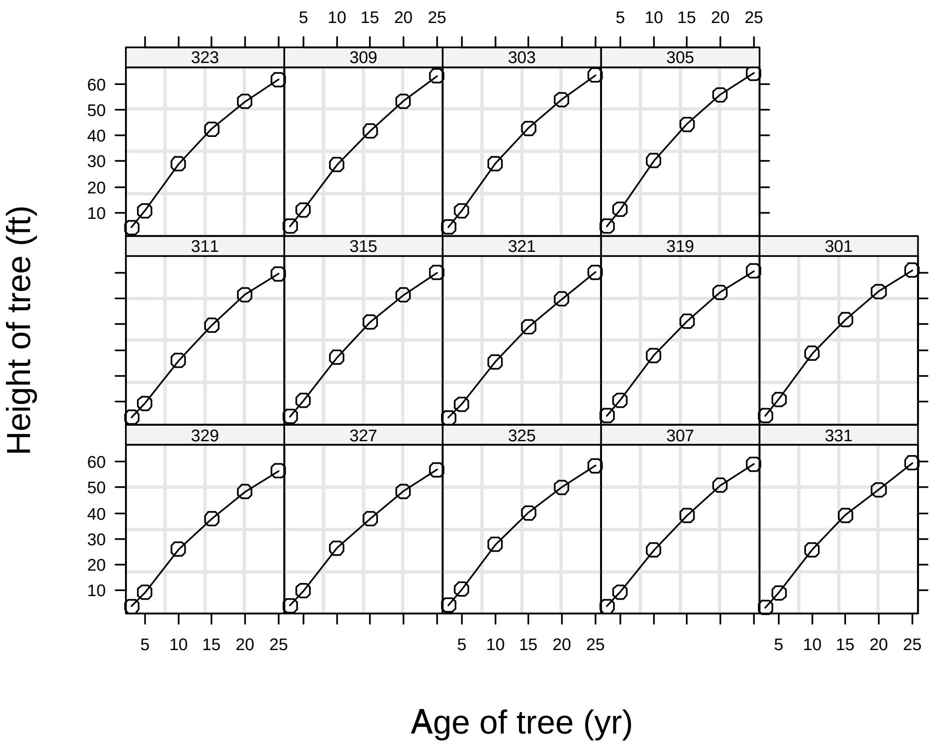

Figure 14: *Height of Loblolly pine trees over time*

based on the columns named:

`height`: height of the tree (ft).

`age`: age of the tree (yr).

`Seed`: a factor indicating the seed source for the tree.

14 Machines — Productivity scores for machines and workers

Data on an experiment to compare three brands of machines used in an industrial process are presented in Milliken and Johnson (1992, §23.1, p. 285). Six workers were randomly chosen among the employees of a factory and operated each machine three times. The response is an overall productivity score taking into account the number and quality of components produced. These data, shown in Figure 15, are analyzed in Milliken and Johnson (1992) with an ANOVA model.

The display formula for these data is

```
score ~ 1 | Worker
```

based on the columns named:

`score`: productivity score.

`Worker`: a factor giving the unique identifier for each worker.

`Machine`: a factor identifying the machine brand - A, B, or C.

15 Meat — Scores for tenderness of beef roasts

Cochran and Cox (1957, §11.51, p. 444) describe data from an experiment conducted at Iowa State College Paul (1943) to compare the effects of length of cold storage on the tenderness of beef roasts. Six storage periods ranging from 0 to 18 days were used. Thirty roasts were scored by four judges on a scale from 0 to 10, with the score increasing with tenderness. The response was the sum of all four scores. Left and right roasts from the same animal were grouped into pairs, which were further grouped into five blocks, according to the muscle from which they were extracted. Different storage periods were applied to each roast within a pair according to a balanced incomplete block design. The data are displayed in Figure 16.

The display formula for these data is

```
score ~ 1 | Pair
```

based on the columns named:

`score`: tenderness score of beef roast

`Pair`: a factor giving the unique identifier for each pair of beef roasts

`Block`: a factor identifying the muscle from which the roast was extracted

`Storage`: an ordered factor specifying the storage treatment - 1 (0 days), 2 (1 day), 3 (2 days), 4 (4 days), 5 (9 days), and 6 (18 days)

16 Milk — Protein content of cows' milk

Diggle, Liang and Zeger (1994) describe data on the protein content of cows' milk in the weeks following calving. The cattle are grouped according to whether they are fed a diet with barley alone, with barley and lupins, or with lupins alone. These data are plotted in Figures 17–19.

The display formula for these data is

```
protein ~ time | Cow
```

based on the columns named:

`protein`: protein content of the milk

`time`: time since calving (weeks)

`Cow`: a factor giving a unique identifier for each cow

`Diet`: a factor identifying the diet as barley, barley+lupins, or lupins

17 Muscle — contraction of heart muscle sections

Baumann and Waldvogel (1963) describe data on the shortening of heart muscle strips dipped in a $CaCl_2$ solution. The muscle strips are taken from the left auricle of a rat's heart. These data, shown in Figure 20 are analyzed in Linder, Chakravarti and Vuagnat (1964).

The display formula for these data is

```
log10(deltal) ~ conc | Strip
```

based on the columns named:

`deltal`: change in the length of the muscle strip (mm).

`conc`: concentration of $CaCl_2$ in which the muscle strip is dipped (mM).

`Strip`: a factor giving a unique identifier for each muscle strip.

Models Linder et al. (1964) used an asymptotic regression model.

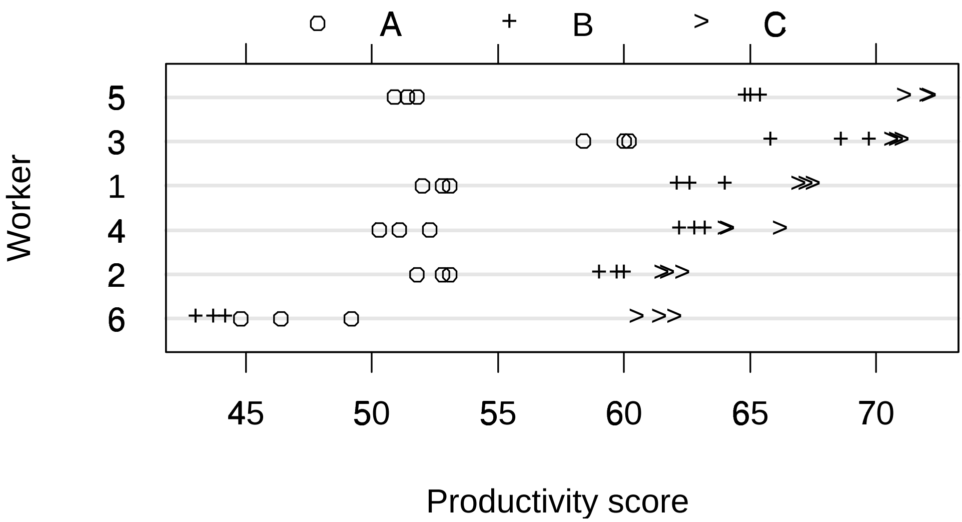

Figure 15: *Productivity scores for three brands of machines. Scores take into acount number and quality of components produced.*

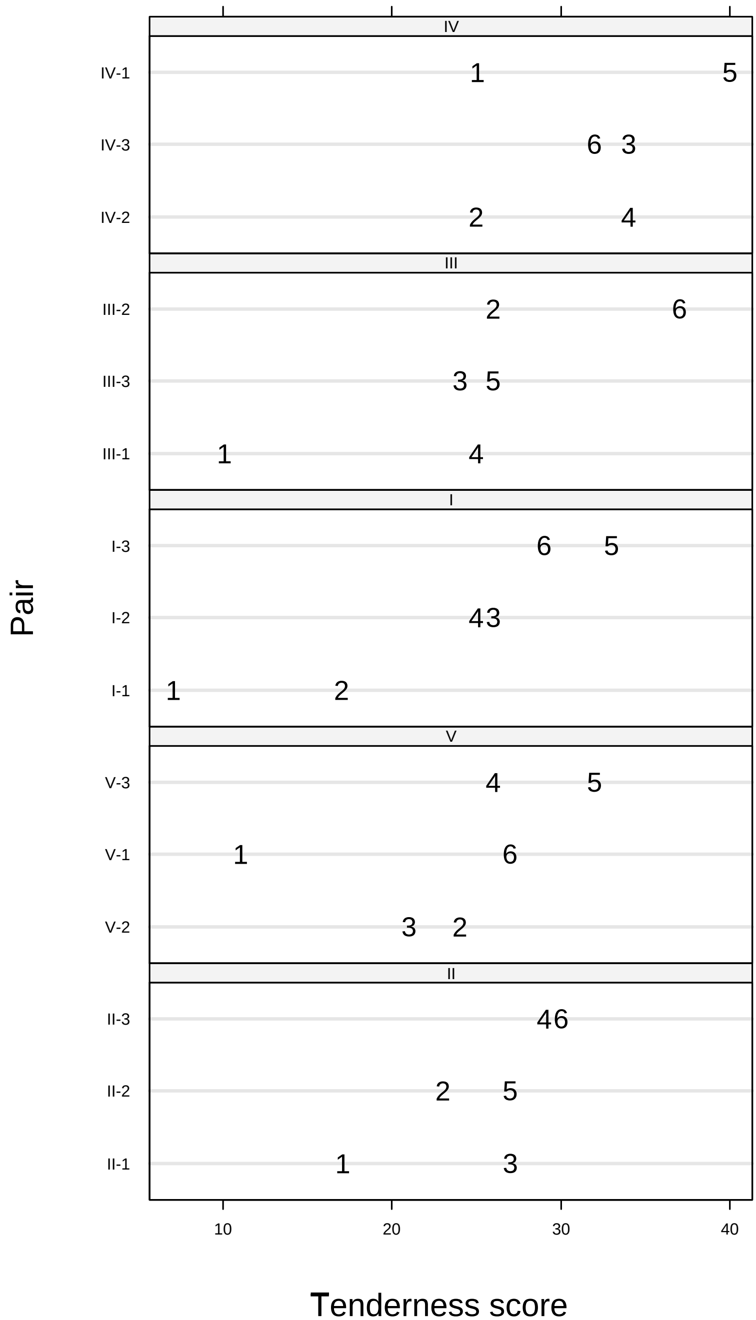

Figure 16: *Tenderness scores of beef roasts stored under cold temperature for different periods of time. Symbols indicate the storage treatment number, which increases with the length of the storage period.*

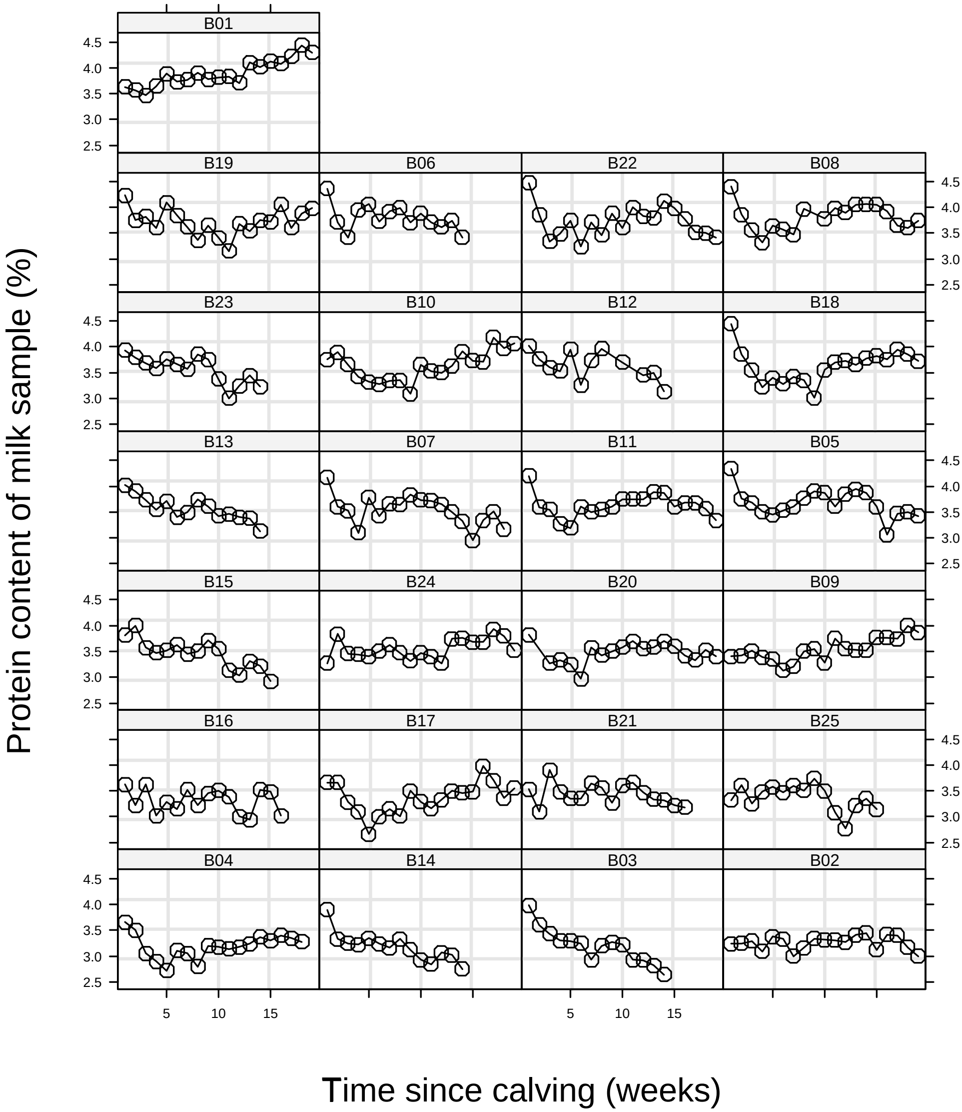

Figure 17: *Protein content of milk versus time since calving — barley diet*

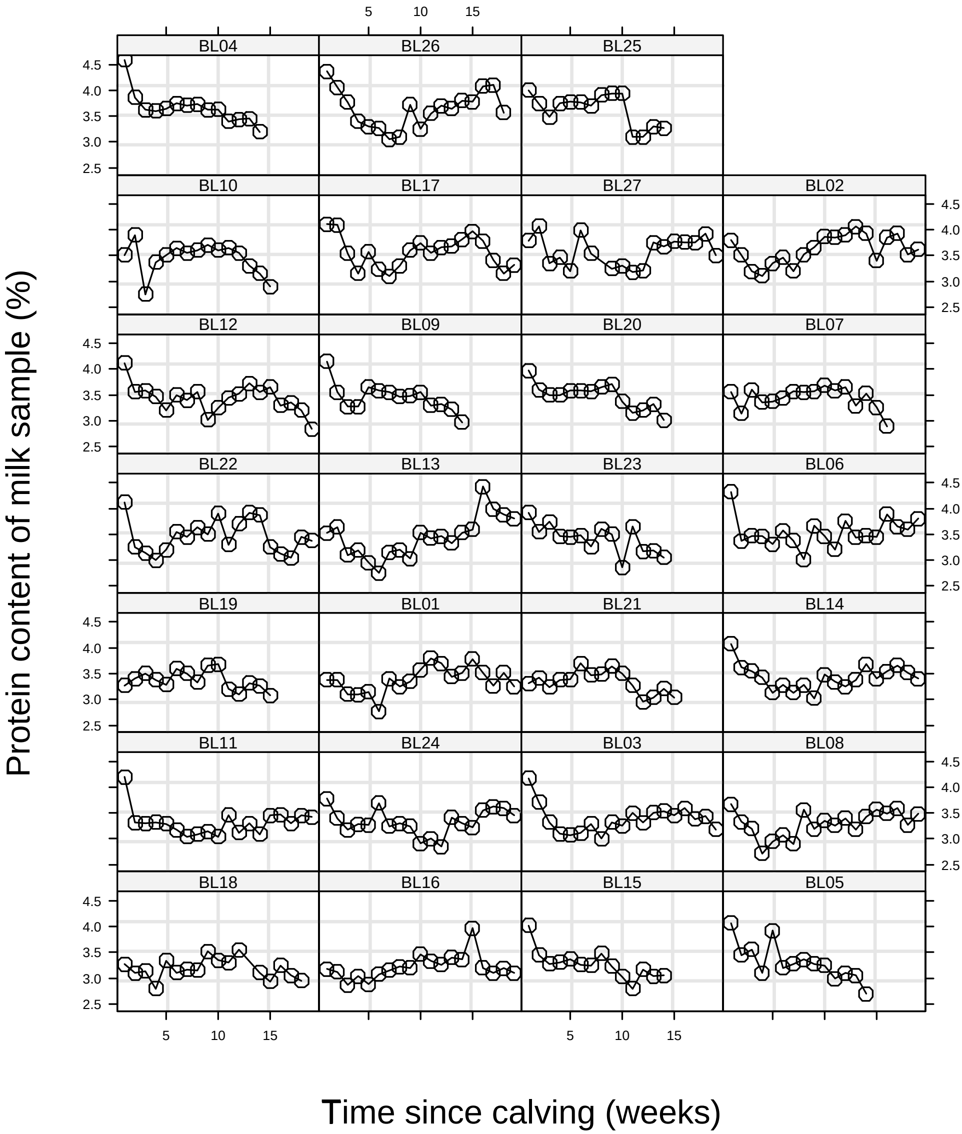

Figure 18: *Protein content of milk versus time since calving — barley and lupins diet*

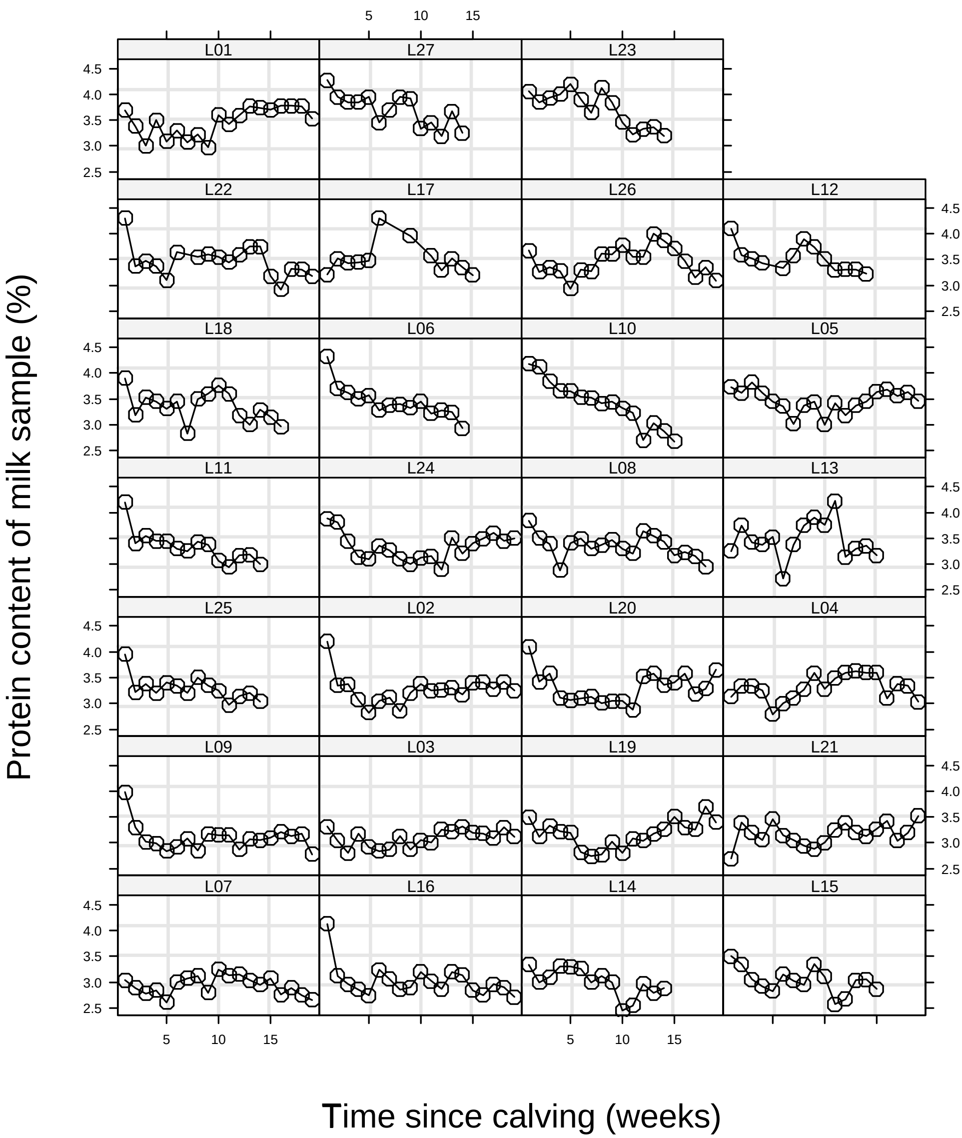

Figure 19: *Protein content of milk versus time since calving — lupins diet*

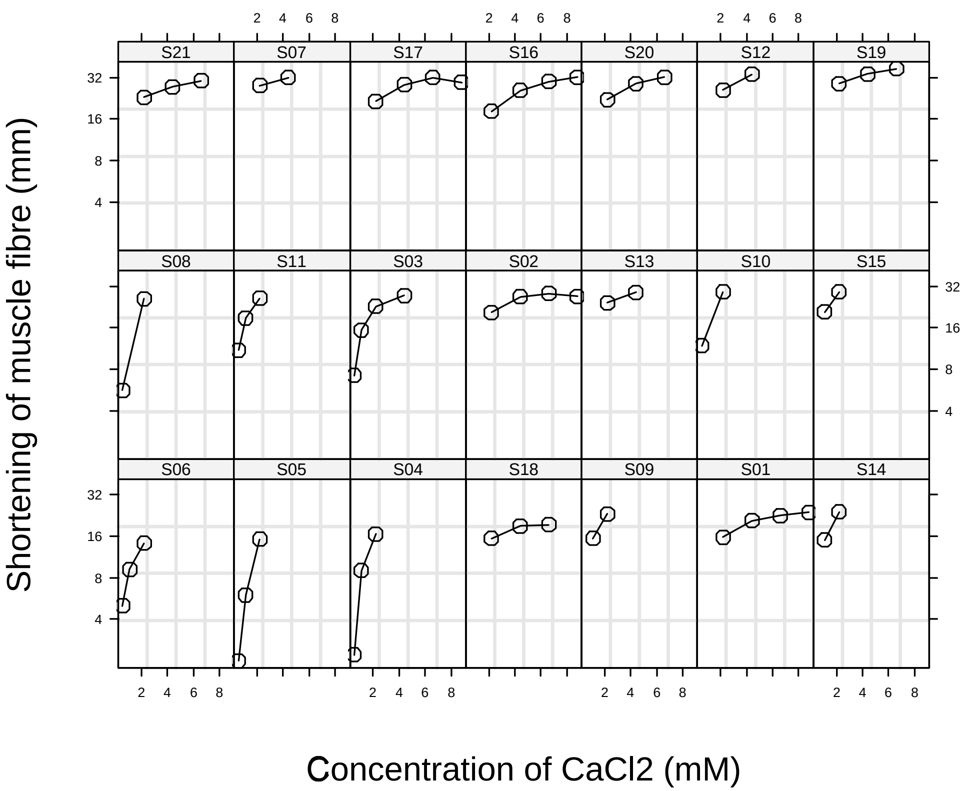

Figure 20: *Logarithm of the change in length of muscle strips from the left auricle of a rat as a function of the concentration of Ca* Cl_2 *in which they are dipped. Twenty-one different strips are used.*

18 Oats — Split-plot experiment on varieties of oats

These data have been introduced by Yates (1935) as an example of a split-plot design. The treatment structure used in the experiment was a 3×4 full factorial, with three varieties of oats and four concentrations of nitrogen. The experimental units were arranged into six blocks, each with three whole-plots subdivided into four sub-plots. One variety of oats was used in each whole-plot with all four concentrations of nitrogen, one concentration in each of the four sub-plots. The data, presented in Figure 21, are analyzed in Venables and Ripley (1996, §6.7, p. 177).

The display formula for these data is

```
yield ~ nitro | Block
```

based on the columns named:

`yield`: the sub-plot yield (bushels/acre).

`Block`: a factor identifying the block - I through VI.

`Variety`: oats variety - Golden Rain, Marvellous, or Victory.

`nitro`: nitrogen concentration (cwt/acre) - 0.0, 0.2, 0.4, 0.6.

19 Orthodont — orthodontic growth data

Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14. Every two years they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head. These data are reported in Potthoff and Roy (1964) and plotted in Figure 22.

The display formula for these data is

```
distance ~ age | Subject
```

based on the columns named:

`distance`: the distance from the center of the pituitary to the pterygomaxillary fissure (mm).

`age`: the age of the subject when the measurement is made (years).

`Subject`: a factor identifying the subject on whom the measurement was made.

`Sex`: a factor indicating if the subject is male or female.

Models: Based on the relationship shown in Figure 22 we begin with a simple linear relationship between `distance` and `age`

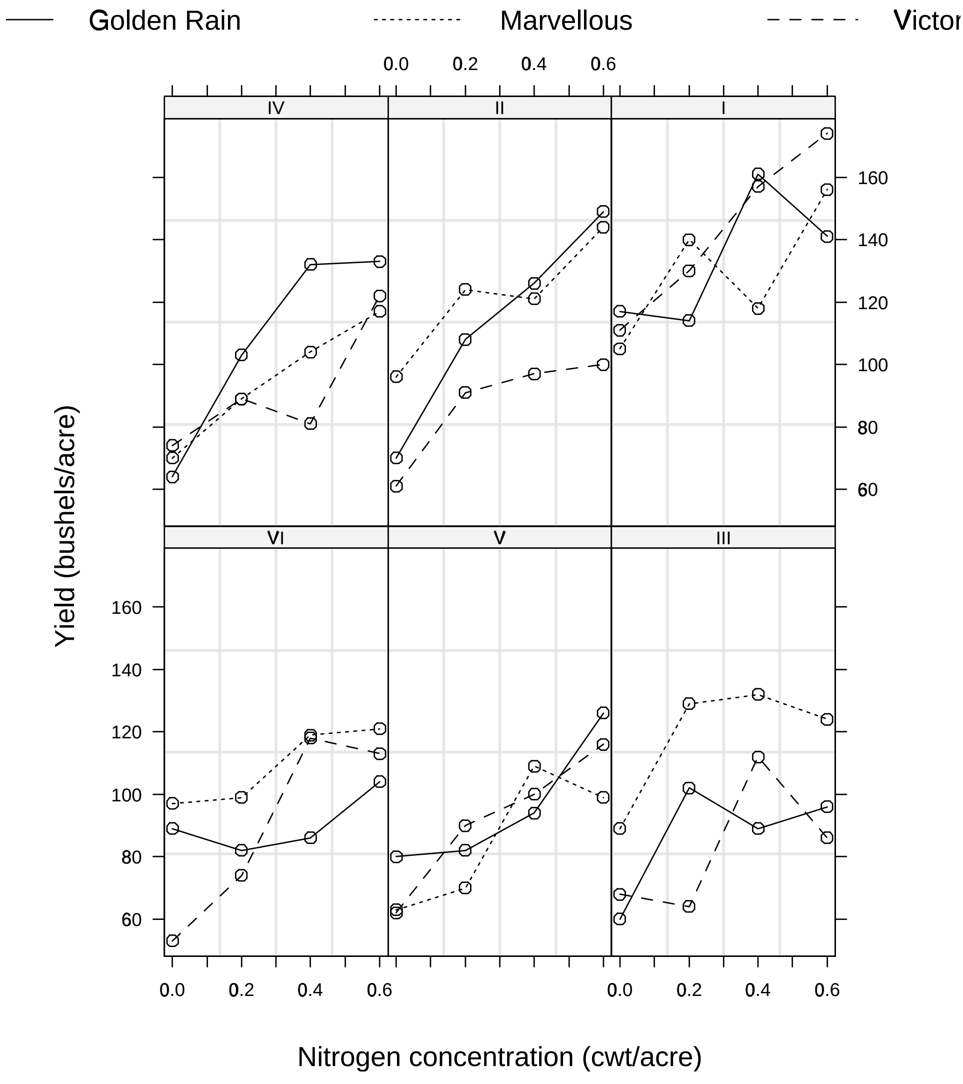

Figure 21: *Sub-plot yields in a split-plot experiment on oats varieties and nitrogen concentrations.*

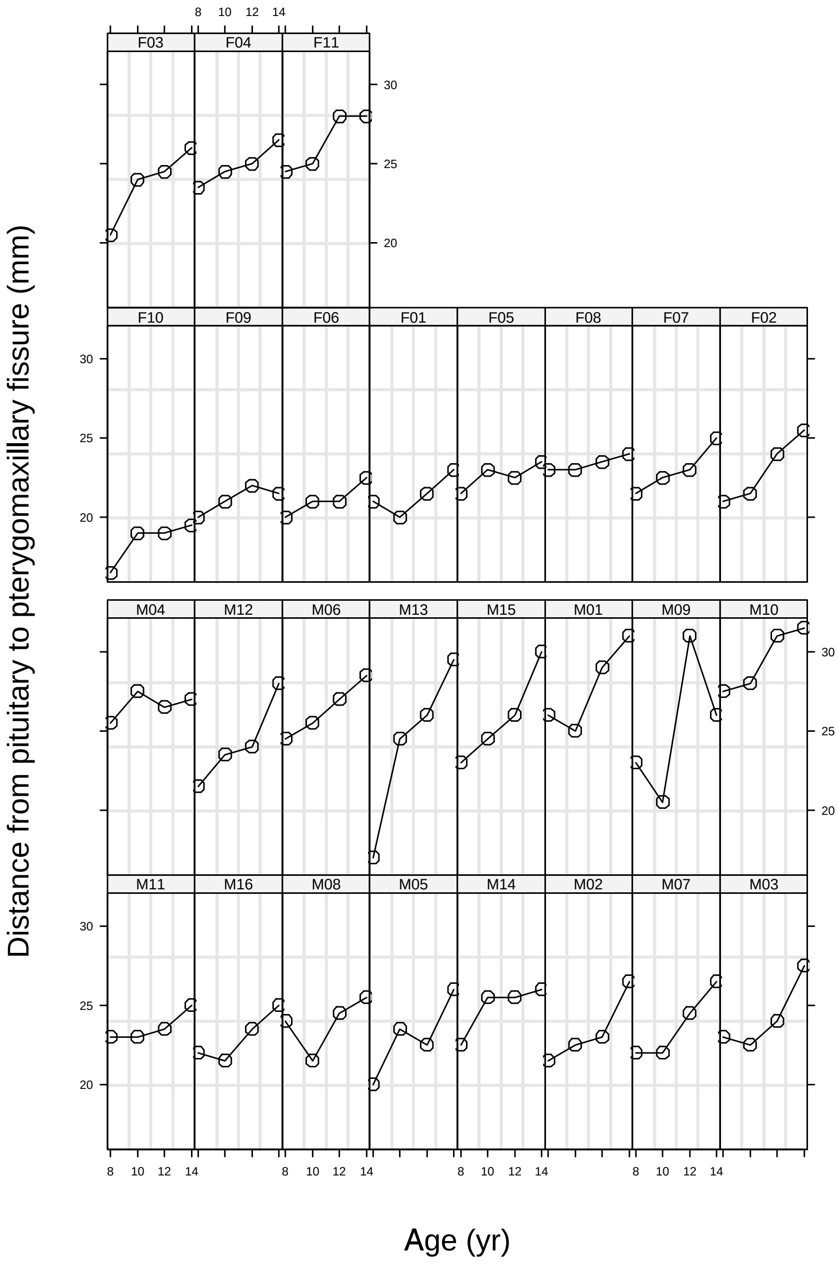

Figure 22: *Distance from the pituitary to the pterygomaxillary fissure versus age for a sample of 16 boys (subjects M01 to M16) and 11 girls (subjects F01 to F11). The aspect ratio for the panels has been chosen to facilitate comparison of the slope of the lines.*

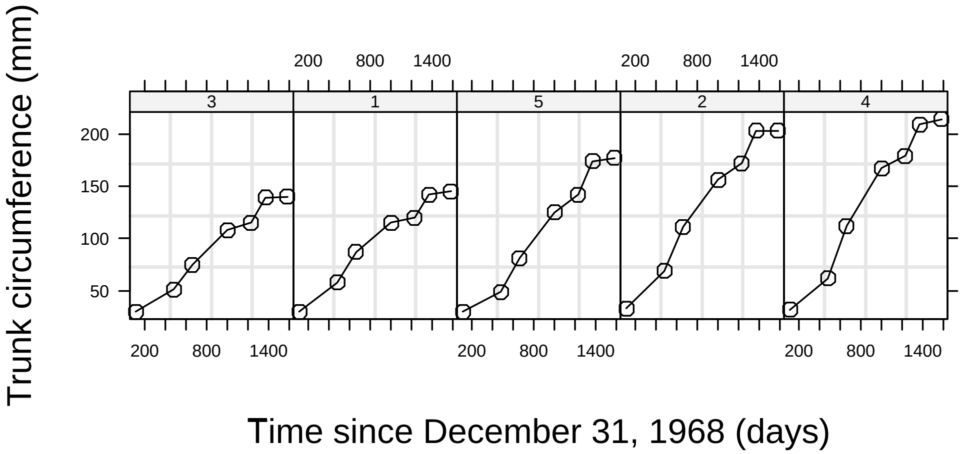

Figure 23: *Circumference of five orange trees from a grove in southern California over time. The measurement is probably the "circumference at breast height" commonly used by foresters.*

20 Orange — Growth of orange trees

Draper and Smith (1981, Exercise 10.N, p. 524) present data on the growth of a group of orange trees. These data are plotted in Figure 23.

The display formula for these data is

```
circumference ~ time | Tree
```

based on the columns named:

`circumference`: circumference of the tree (mm)

`time`: time in days past the arbitrary origin of December 31, 1968.

`Tree`: a factor identifying the tree on which the measurement is made.

Models The logistic growth model provides a reasonable fit to these data.

21 Ovary - Counts of ovarian follicles

Pierson and Ginther (1987) report on a study of the number of large ovarian follicles detected in different mares at several times in their estrus cycles. These data are shown in Figure 24.

The display formula for these data is

```
follicles ~ time | Mare
```

based on the columns named:

`follicles`: the number of ovarian follicles greater than 10 mm. in diameter.

`time`: time in the estrus cycle. The data were recorded daily from three days before ovulation until three days after the next ovulation. The measurement times for each mare are scaled so that the ovulations for each mare occur at times 0 and 1.

`Mare`: a factor indicating the mare on which the measurement is made.

Models:

```
Ovary.lis <- lmList(follicles ~ I(sin(2*pi*time)) + I(cos(2*pi*time),
                   data = Ovary, cluster = ~ Mare)
Ovary.lme <- lme(Ovary.lis)
```

22 Oxide — Variability in semiconductor manufacturing

These data are described in Littell et al. (1996, § 4.4, p. 155) as coming "from a passive data collection study in the semiconductor industry where the objective is to estimate the variance components to determine the assignable causes of the observed variability." The observed response is the thickness of the oxide layer on silicon wafers, measured at three different sites of each of three wafers selected from each of eight lots sampled from the population of lots. We display the data in Figure 25

The display formula for these data is

```
Thickness ~ 1 | Lot/Wafer
```

based on the columns named:

`Thickness`: thickness of the oxide layer.

`Lot`: a factor giving a unique identifier for each lot.

`Wafer`: a factor giving a unique identifier for each wafer within a lot.

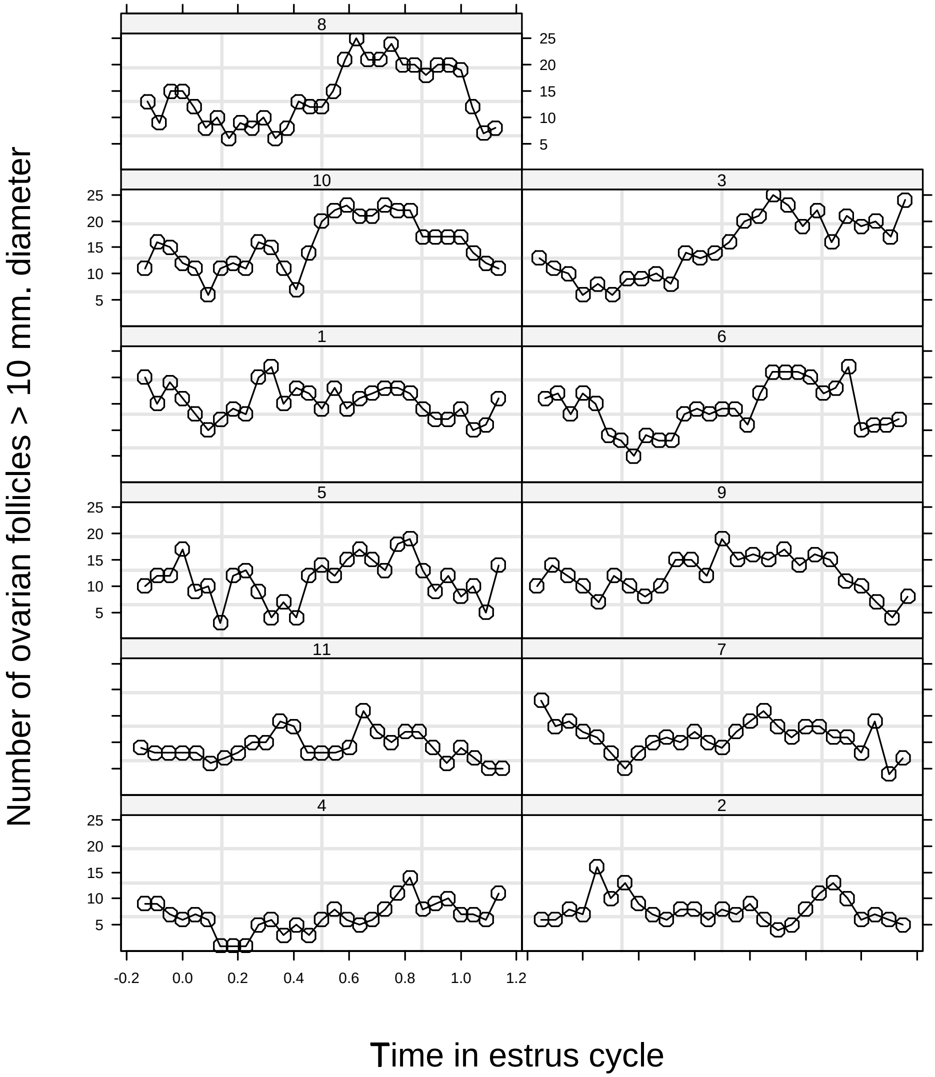

Figure 24: *Number of ovarian follicles greater than 10 mm in diameter detected in mares at various times in their estrus cycles. The times have been scaled so the ovulations occur at times 0 and 1.*

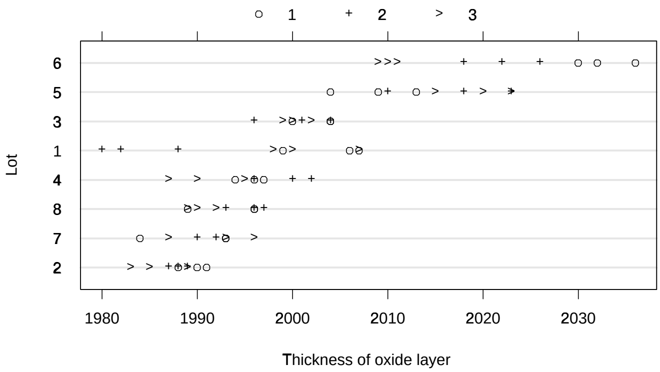

Figure 25: *Thickness of oxide layer measured on different sites of wafers selected from a sample of manufacturing lots. Symbols denote different wafers within the same lot.*

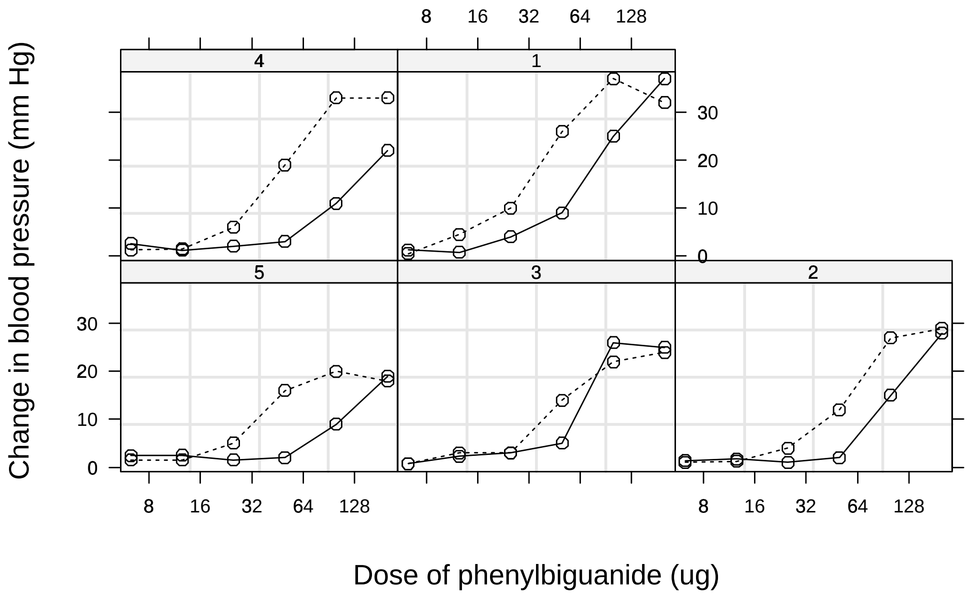

Figure 26: *Change in blood pressure versus dose of phenylbiguanide (PBG) for five rabbits. Each rabbit was exposed to increasing doses of PBG once after treatment with the H_5^3-agonist MDL 72222 and once after treatment with a placebo.*

23 PBG — effect of phenylbiguanide on blood pressure

Data on an experiment to examine the effect of a antagonist MDL 72222 on the change in blood pressure experienced with increasing dosage of phenylbiguanide are described in Ludbrook (1994) and analysed in Venables and Ripley (1996). Each of five rabbits was exposed to increasing doses of phenylbiguanide after having either a placebo or the HD_5-antagonist MDL 72222 administered. The data are shown in Figure 26.

The display formula for these data is

```
deltaBP ~ PBG | Rabbit
```

based on the columns named:

`deltaBP`: change in blood pressure (mm Hg).

`PBG`: dose of phenylbiguanide (μg).

`Rabbit`: a factor identifying the test animal.

`Treatment`: a factor identifying whether the observation was made after administration of `placebo` or the HD_5-antagonist `MDL 72222`.

Models The form of the response suggests a logistic model for the change in blood pressure as function of the logarithm of the concentration of PBG.

24 Phenobarb — Phenobarbital kinetics

Data from a pharmacokinetics study of phenobarbital in neonatal infants. During the first few days of life the infants receive multiple doses of phenobarbital for prevention of seizures. At irregular intervals blood samples are drawn and serum phenobarbital concentrations are determined. The data were originally given in Grasela and Donn (1985). They are also analysed in Boeckmann, Sheiner and Beal (1994) and in Davidian and Giltinan (1995, § 6.6).

The display formula for these data is

```
conc ~ time | Subject
```

based on the columns named:

`conc`: phenobarbital concentration in the serum (ug/L).

`time`: time when the sample is drawn or drug administered (hr).

`Subject`: a factor identifying the infant.

`Wt`: birth weight of the infant (kg).

`Apgar`: the 5-minute Apgar score for the infant. This is an indication of health of the newborn infant. The scale is 1 to 10.

`dose`: dose of drug administered (ug/kg).

25 Pixel — Pixel intensity in lymphnodes

These data are from an experiment conducted by Deborah Darien, Department of Medical Sciences, School of Veterinary Medicine, University of Wisconsin - Madison. The mean pixel intensity of the right and left lymphnodes in the axillary region obtained from CT scans of 10 dogs were recorded over a period of 14 days after intravenous application of a contrast. The data are shown in Figure 27.

The display formula for these data is

```
pixel ~ day | Dog
```

based on the columns named:

`pixel`: mean pixel intensity of lymphnode in the CT scan.

`day`: number of days since contrast administration.

`Dog`: a factor giving the unique identifier for each dog.

`Side`: a factor indicating the side on which the measurement was made.

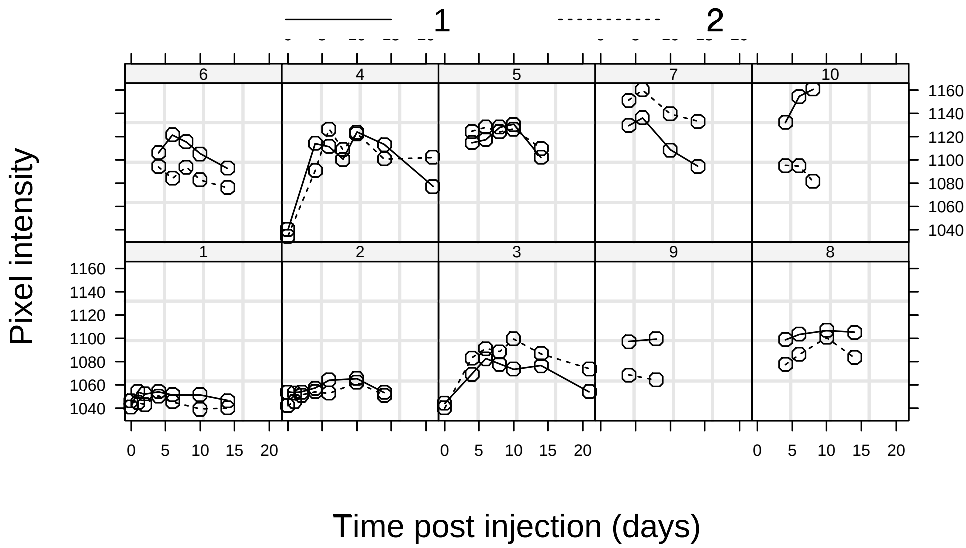

Figure 27: *Mean pixel intensity of the right and left lymphnodes in the axillary region versus time from intravenous application of a contrast. The pixel intensities were obtained from CT scans.*

26 Quinidine — Quinidine kinetics

Verme, Ludden, Celmenti and Harris (1992) analyze routine clinical data on patients receiving the drug quinidine as a treatment for cardiac arrythmia (atrial fibrillation of ventricular arrythmias). All patients were receiving oral quinidine doses. At irregular intervals blood samples were drawn and serum concentrations of quinidine were determined. These data are analysed in several places including Davidian and Giltinan (1995, § 9.3).

The display formula for these data is

```
conc ~ time | Subject
```

based on the columns named:

`conc`: Serum quinidine concentration (mg/L).

`time`: Time (hr.) at which the drug was administered or the blood sample drawn. This is measured from the time the patient entered the study.

`Subject`: A factor identifying the patient on whom the data were collected.

`dose`: Dose of drug administered (mg). Although there were two different forms of quinidine administered, the doses were adjusted for differences in salt content by conversion to milligrams of quinidine base.

`interval`: When the drug has been given at regular intervals for a sufficiently long period of time to assume steady state behaviour, the interval is recorded.

`Age`: Age of the subject on entry to the study (yr).

`Height`: Height of the subject on entry to the study (in).

`Weight`: Body weight of the subject (kg).

`Race`: A factor with possible three levels: `Caucasian`, `Black`, `Latin`.

`Smoke`: A factor giving smoking status at the time of the measurement: `no` or `yes`.

`Ethanol`: A factor giving ethanol (alcohol) abuse status at the time of the measurement: `none`, `current`, or `former`.

`Heart`: A factor indicating congestive heart failure for the subject: `none or mild`, `moderate`, or `severe`.

`Creatinine`: A factor in 8 levels coding the creatinine clearance and other measurements. Creatinine clearance is divided into those greater than 50 mg/min and those less than 50 mg/min.

`glyco`: Alpha-1 acid glycoprotein concentration (mg/dL). Often measured at the same time as the quinidine concentration.

27 Rail — Evaluation of stress in rails

Devore (1995, Example 10.10, p. 414) cites data from an article in *Materials Evaluation* on "a study to travel time for a certain type of wave that results from longitudinal stress of rails used for railroad track".

The display formula for these data is

```
travel ~ 1 | Rail
```

based on the columns named:

`travel`: Travel time for ultrasonic head-waves in the rail (nanoseconds). The value given is the original travel time minus 36100 nanoseconds.

`Rail`: a factor giving the number of the rail on which the measurement was made.

28 Relaxin — Bioassay of the Protein Relaxin

Fei, Gross, Lofgren, Mora-Worms and Chen (1990) present data on nine runs of a bioassay of the protein relaxin. These data also are described in Davidian and Giltinan (1995, § 1.1.4) and analyzed in their section 10.4.2 .

The display formula for these data is

```
cAMP ~ conc | Run
```

based on the columns named:

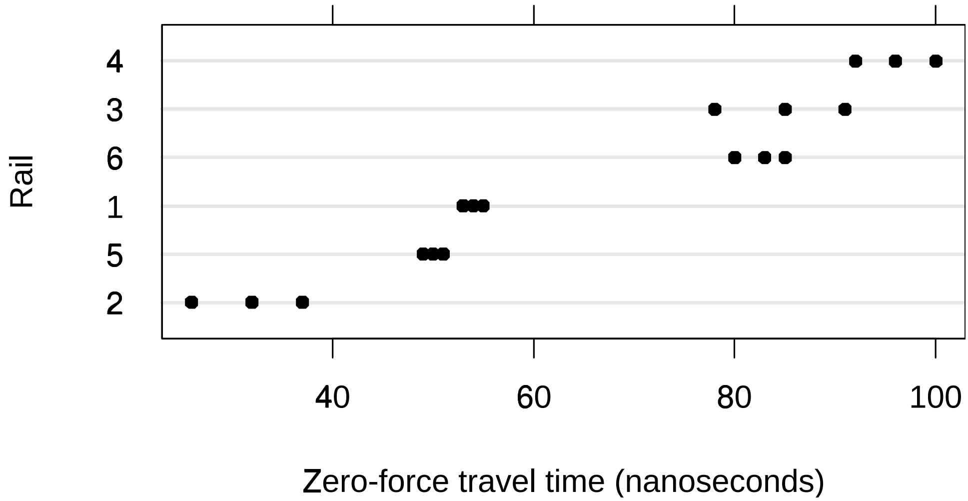

Figure 28: *Travel time for ultrasonic head-waves in a sample of six railroad rails. The travel times shown here in nanoseconds are the result of subtracting 36.1 μs from the original observations. That is, the original observation would be 36100 nanoseconds plus the value shown here. The vertical axis indicates the rail on which the measurements were made.*

`cAMP`: amount of intracellular adenosine-3',5'-cyclic monophosphate (cAMP) released (pmole/ml).

`conc`: known concentration of the relaxin protein (ng/ml).

`Run`: a factor giving the run number.

Models Davidian and Giltinan (1995) use the four-parameter logistic model with these data.

29 Soybean — soybean leaf weight over time

These data are described in Davidian and Giltinan (1995, § 1.1.3, p. 7) as "Data from an experiment to compare growth patterns of two genotypes of soybeans: Plant Introduction #416937 (P), an experimental strain, and Forrest (F), a commercial variety." We display them in Figures 30 and 31

The display formula for these data is

```
weight ~ time | Plot
```

based on the columns named:

`weight`: average leaf weight per plant (g).

`time`: time the sample was taken (days after planting).

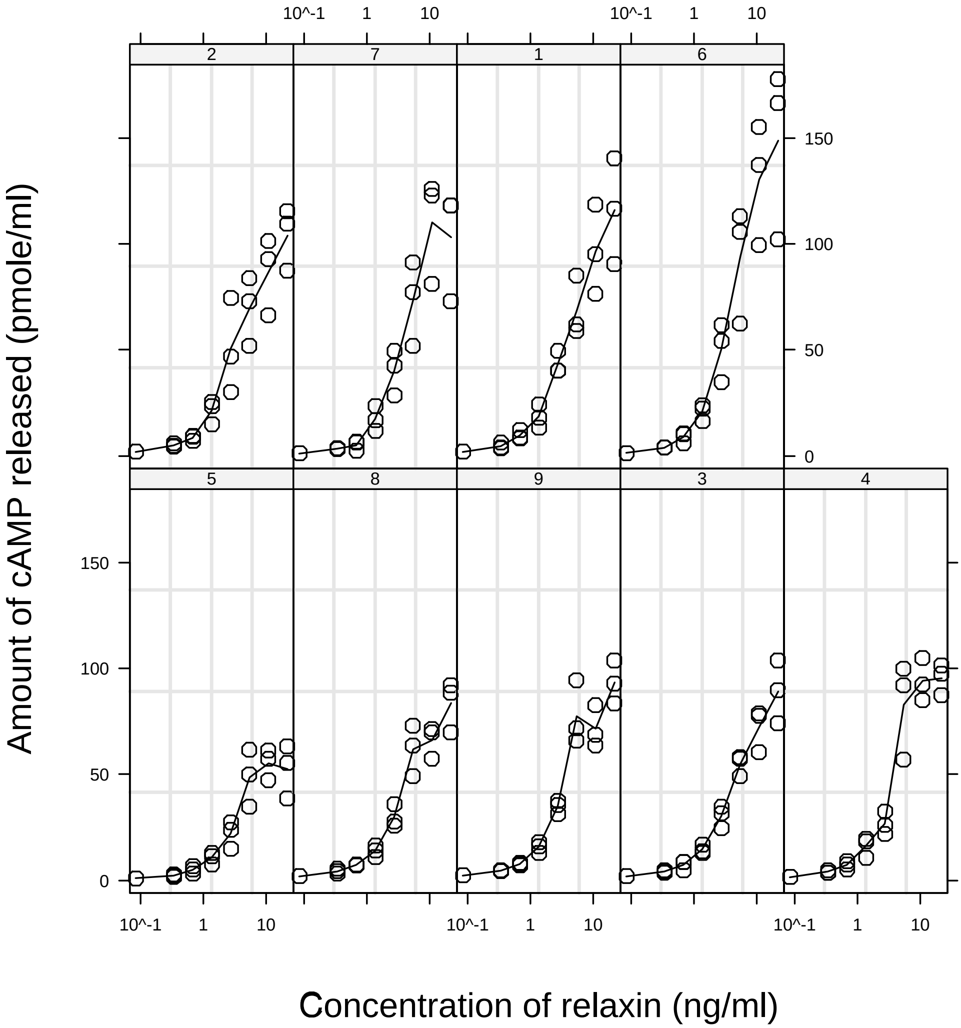

Figure 29: *Amount of cAMP released versus concentration of relaxin for nine different runs of a bioassay of the protein.*

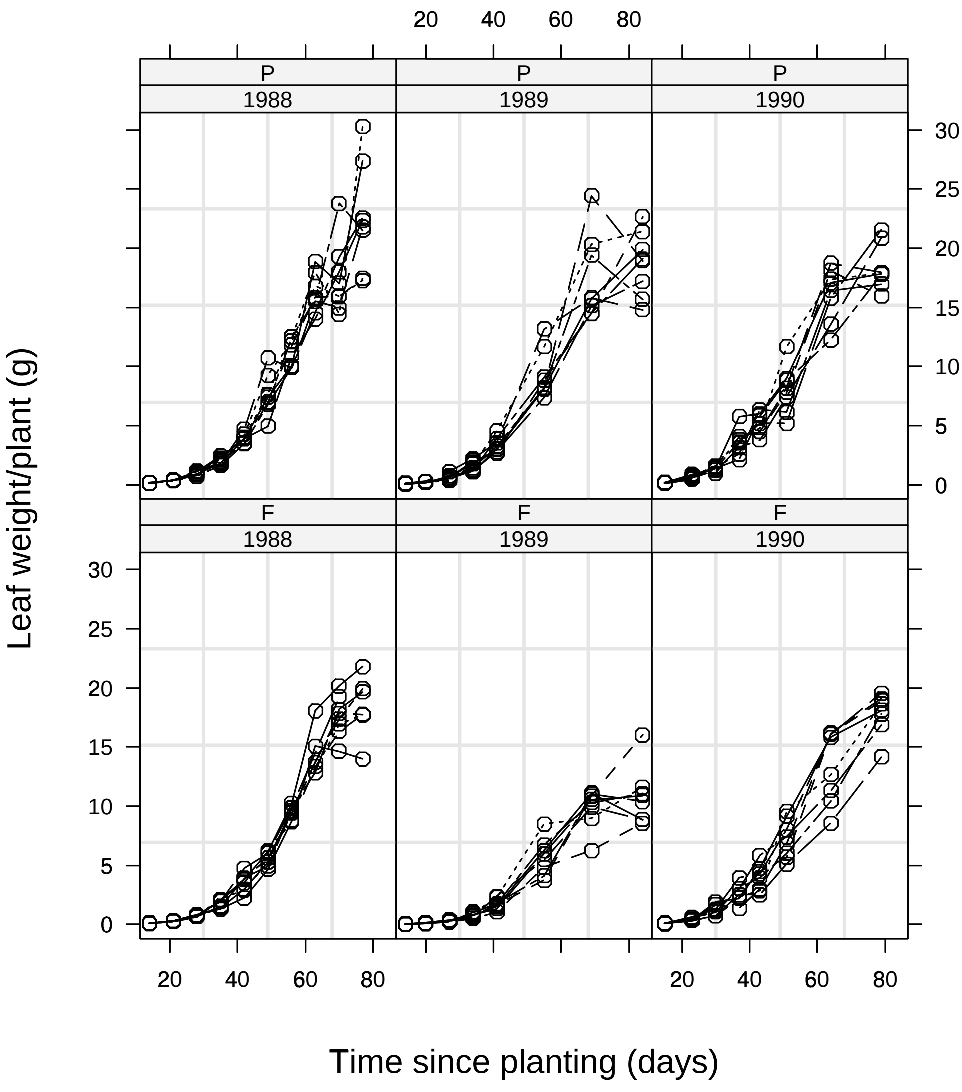

Figure 30: *Average leaf weight per plant versus time since planting for plots of soybeans. The plots are from two different years and represent two different genotypes of soybeans.*

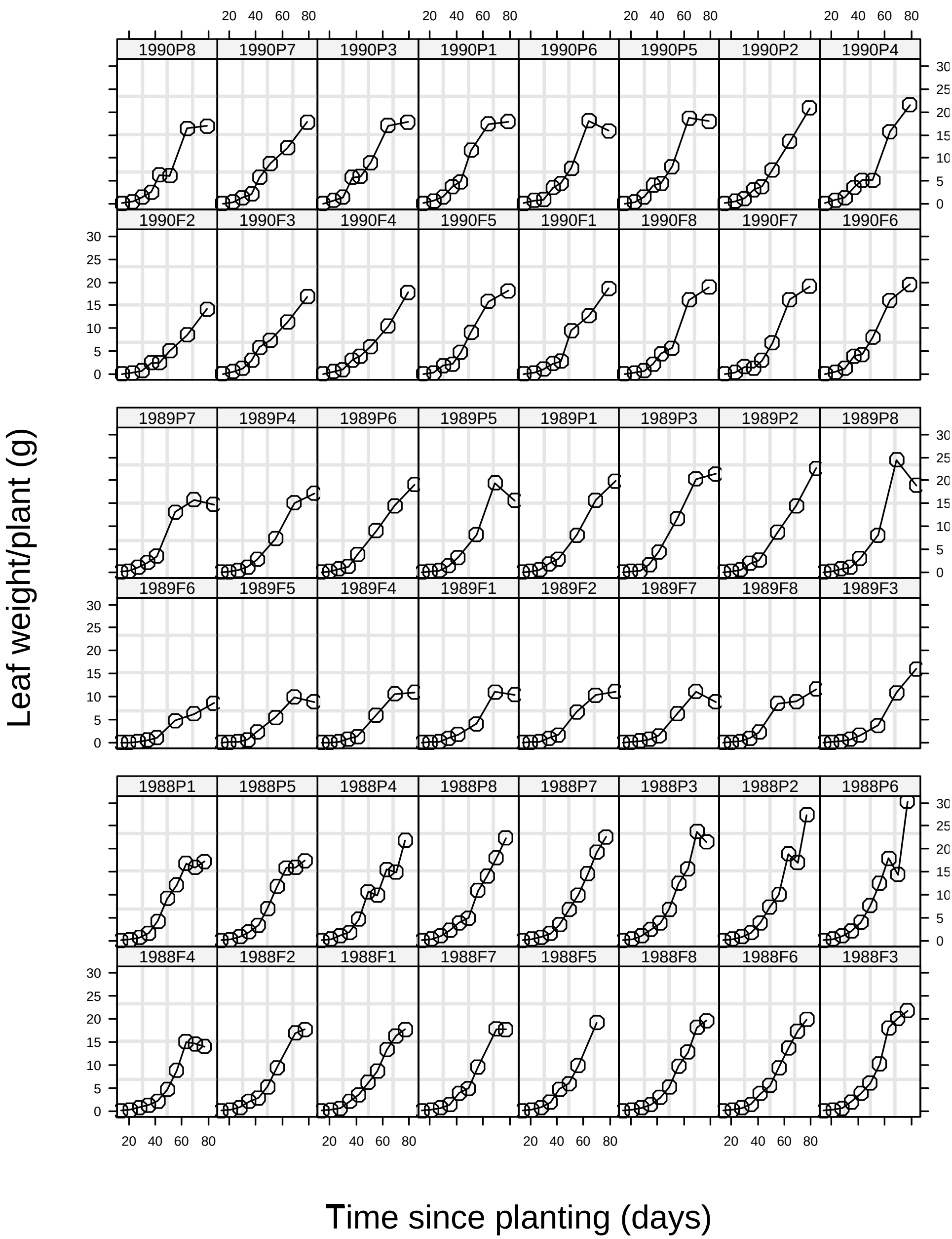

Figure 31: *Average leaf weight per plant versus time since planting for plots of soybeans. The plots are from two different years and represent two different genotypes of soybeans.*

`Plot`: a factor giving a unique identifier for each plot.

`Variety`: a factor indicating variety F or variety P for the plot.

`Year`: the year the plot was planted.

Models Potvin et al. (1990) suggest using a modified form of the asymptotic regression model.

30 Spruce — Growth of spruce trees

Diggle et al. (1994, Example 1.3, page 5) describe data on the growth of spruce trees that have been exposed to an ozone-rich atmosphere or to a normal atmosphere. These data are plotted in Figures 32–34.

The display formula for these data is

```
logSize ~ days | Tree
```

based on the columns named:

`logSize`: the logarithm of an estimate of the volume of the tree trunk

`days`: number of days since the beginning of the experiment

`Tree`: a factor giving a unique identifier for each tree

`Plot`: a factor identifying the plot in which the tree was grown. The levels of this factor are `Ozone1`, `Ozone2`, `Normal1`, and `Normal2`.

`Treatment` a factor indicating whether the tree was grown in an ozone-rich atmosphere or a normal atmosphere.

31 Theophylline — Theophylline kinetics

Boeckmann et al. (1994) report data from a study by Dr. Robert Upton of the kinetics of the anti-asthmatic drug Theophylline. Twelve subjects were given oral doses of Theophylline then serum concentrations were measured at 11 time points over the next 25 hours. Davidian and Giltinan (1995) also analyse these data, shown in Figures 35 and 36.

The display formula for these data is

```
conc ~ time | Subject
```

based on the columns named:

`conc`: Theophylline concentration in the sample (mg/L).

`time`: time since drug administration when the sample was drawn (hours).

`Subject`: a factor identifying the subject.

`Wt`: weight of the subject (kg).

`Dose`: dose administered to the subject (mg/kg).

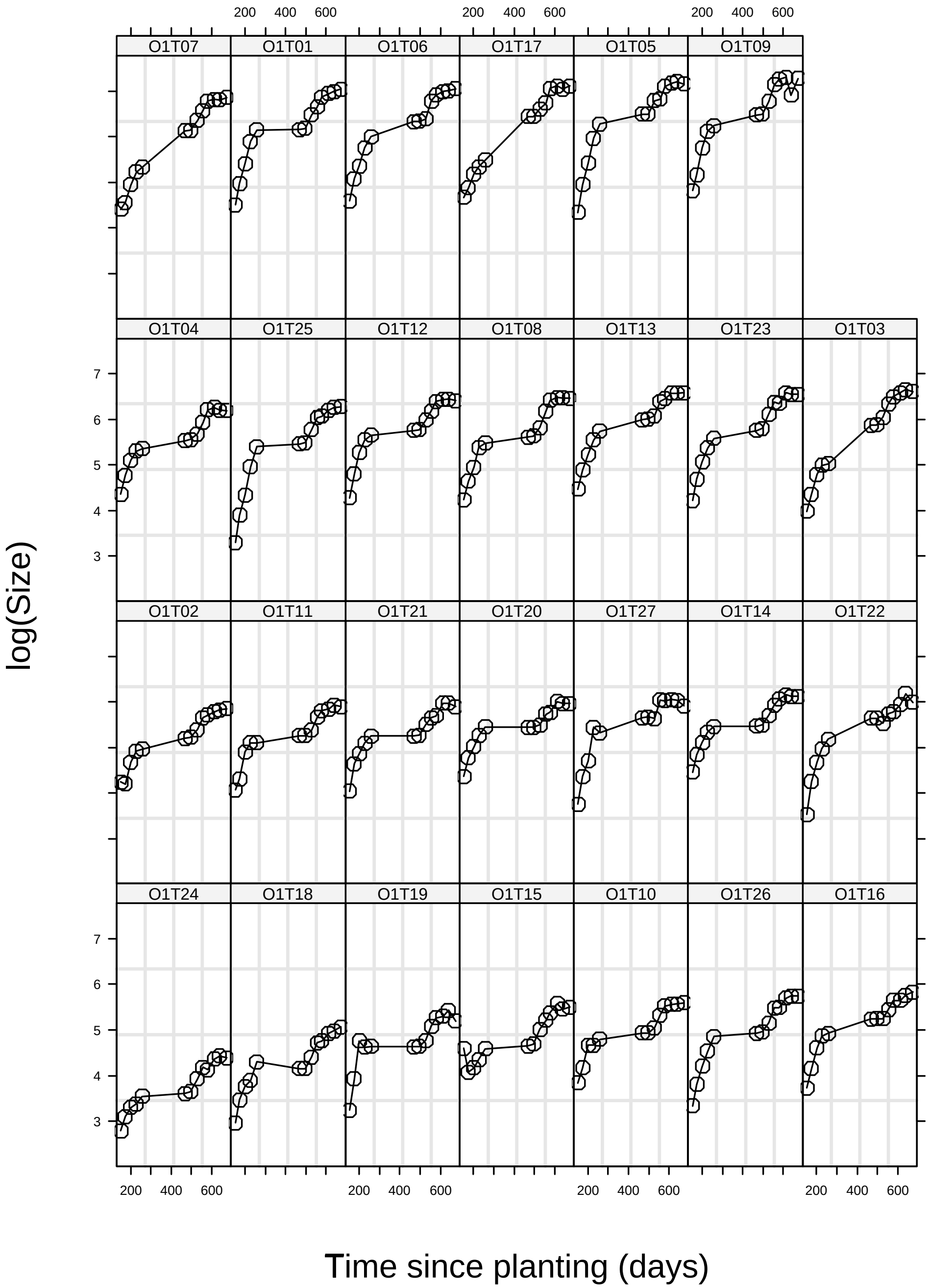

Figure 32: *Growth measures in the logarithm of an estimate of the volume of the spruce tree trunk versus time. These 27 trees were in the first plot that was exposed to an ozone-rich atmosphere throughout the experiment*

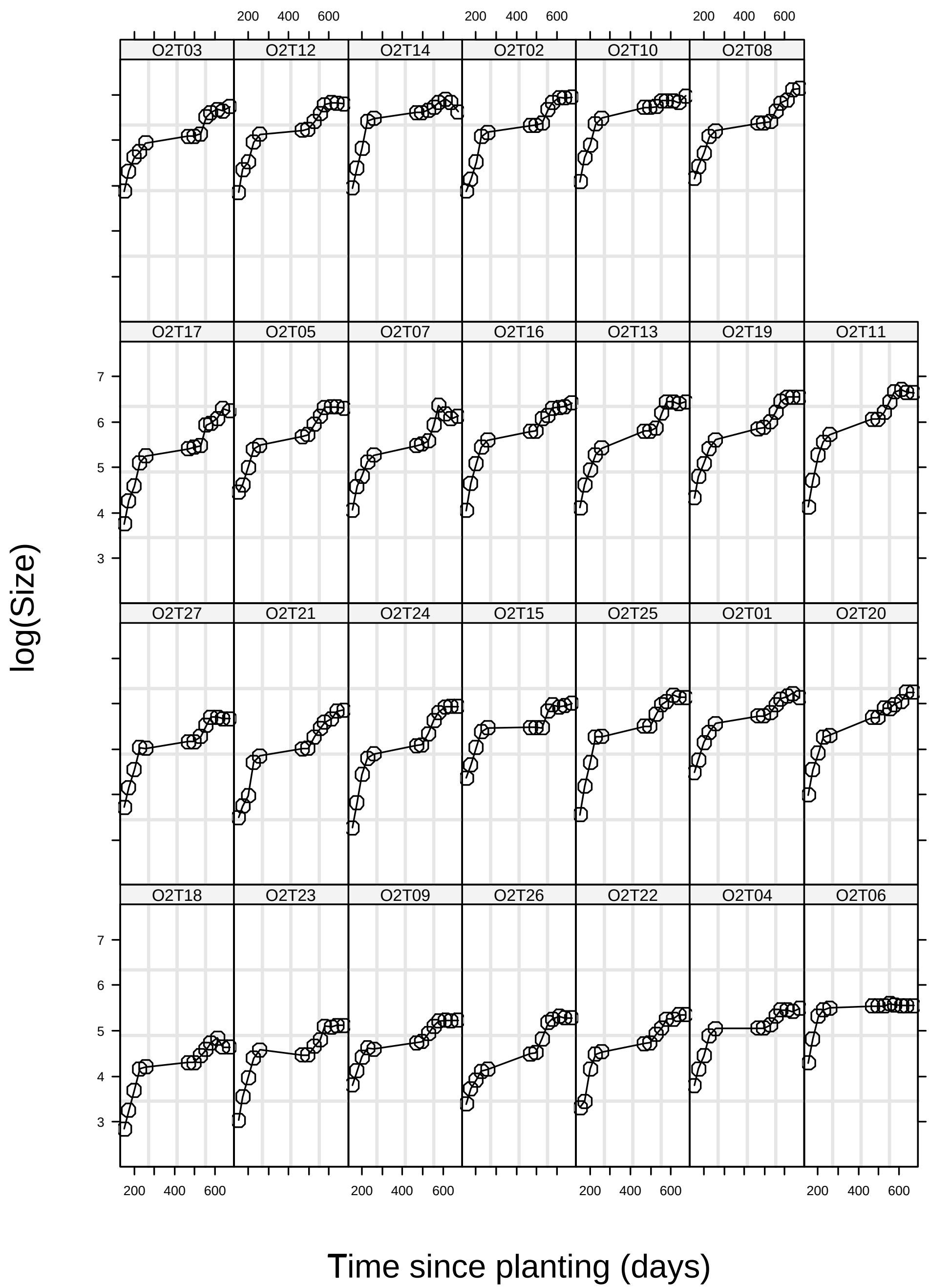

Figure 33: *Growth measures in the logarithm of an estimate of the volume of the spruce tree trunk versus time. These 27 trees were in the second plot that was exposed to an ozone-rich atmosphere throughout the experiment*

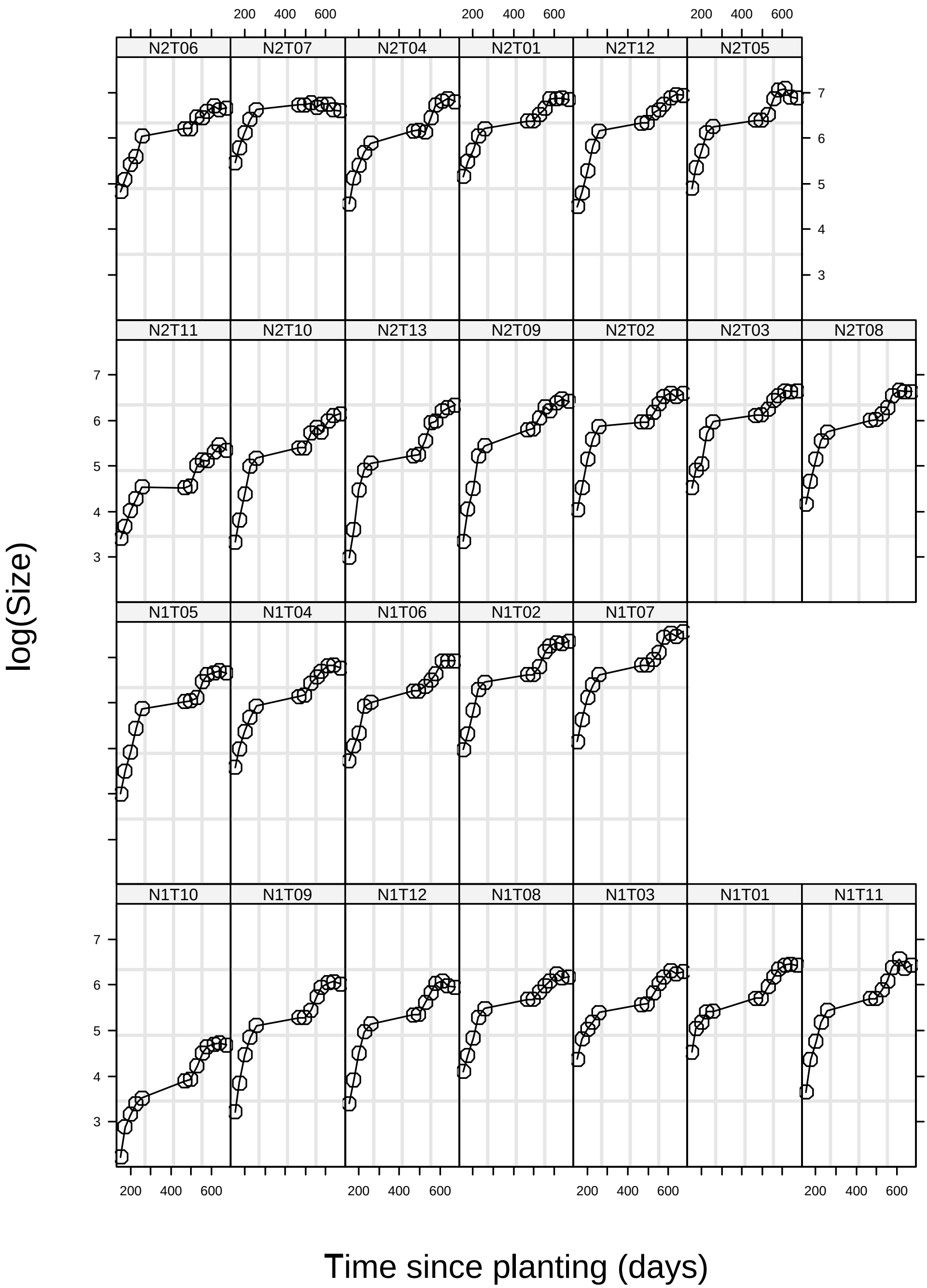

Figure 34: *Growth measures in the logarithm of an estimate of the volume of the spruce tree trunk versus time. These 25 trees were in the first and second plot that were exposed to an normal atmosphere throughout the experiment*

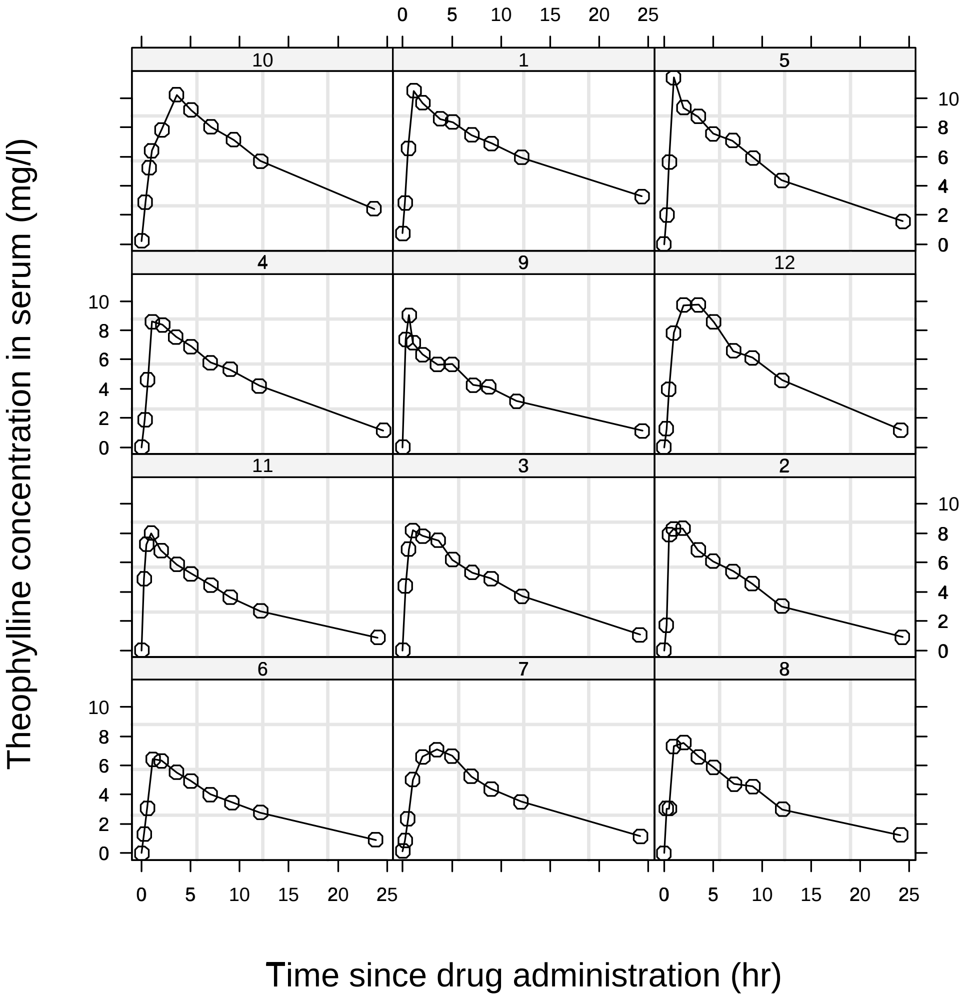

Figure 35: *Serum concentrations of Theophylline versus time since oral administration of the drug in twelve subjects.*

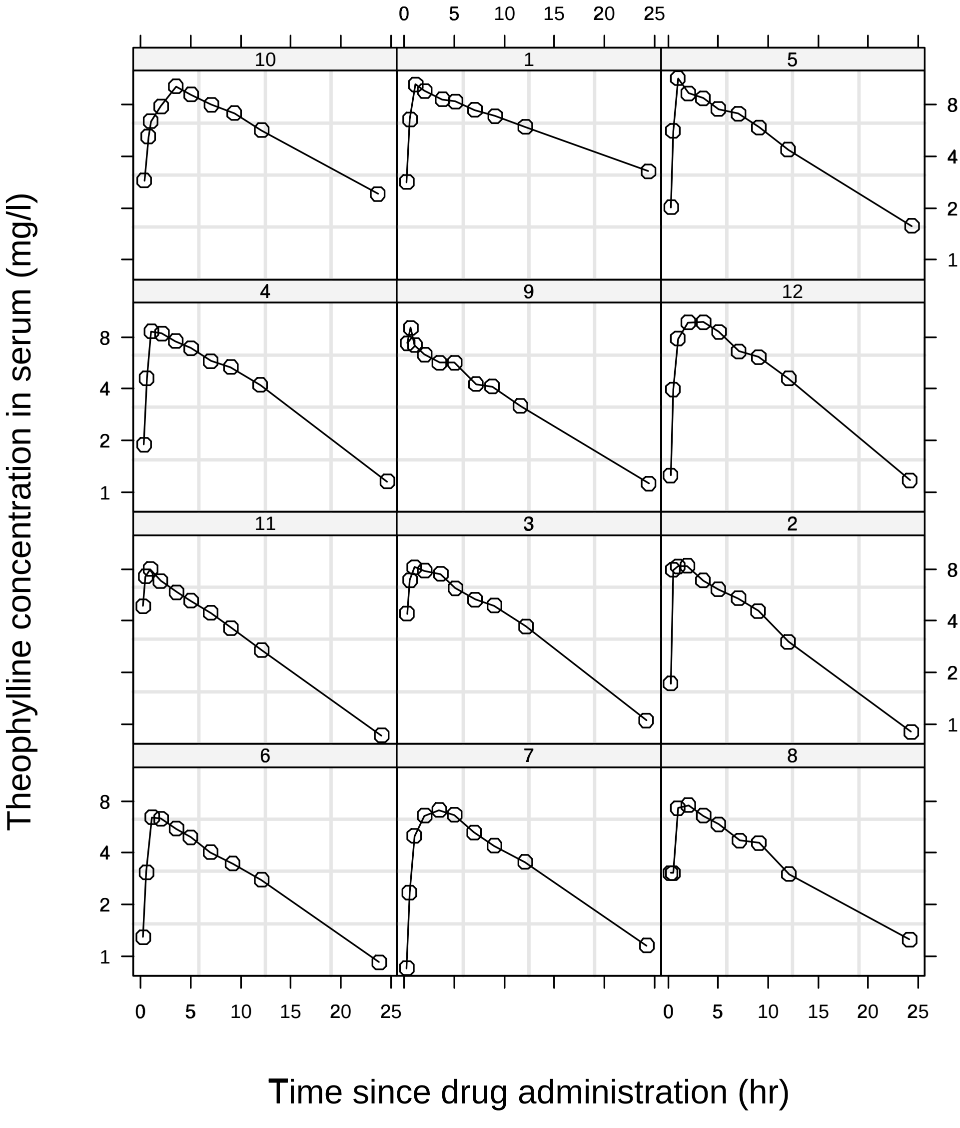

Figure 36: *Serum concentrations of Theophylline versus time since oral administration of the drug in twelve subjects. The serum concentrations are represented on a logarithmic scale. The observations at* time $= 0$ *are removed.*

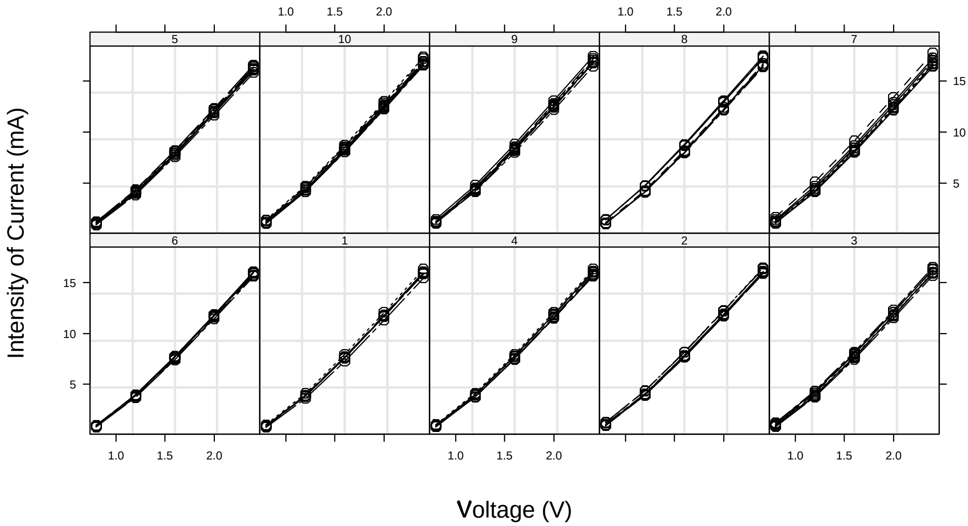

Figure 37: *Current versus voltage curves for each site, by wafer.*

Models: Both Boeckmann et al. (1994) and Davidian and Giltinan (1995) use a two-compartment open pharmacokinetic model for these data.

32 Wafer — Modeling of analog circuits

In an experiment to study the variability in the manufacturing of analog MOS circuits, the intensity of current (in Amperes) at five ascending voltages (in Volts) were collected on n-channel devices. Measurements were made on 8 sites of each of 10 wafers. Figure 37 shows the response curves for each site, by wafer.

The display formula for these data is

```
current ~ voltage | Wafer/Site
```

based on the columns named:

`current`: the intensity of current (mA).

`voltage`: the voltage applied to the device (V).

`Wafer`: a factor giving a unique identifier for each wafer.

`Site`: a factor giving an identifier for each site within a wafer.

33 Wheat — Wheat dry matter amount

Milliken and Johnson (1992, §24.4, p. 308) present data on an experiment “where the dry matter amount was measured on wheat plants grown under different levels of moisture and different amounts of fertilizer." Four equally spaced levels of moisture (10, 20, 30, and 40 ml) and four equally spaced amounts of fertilizer (2, 4, 6, and 8 mg) were used in a full factorial arrangement. The 48 peat pots available for the experiment were arranged in groups of 4 and put into 12 trays. The treatment combinations were assigned to the experimental units according to a split-plot design, with moisture as the whole-plot treatment (applied to trays) and fertilizer as the sub-plot treatment (applied to pots within trays). The response was the dry matter amount 30 days after planting. The data are presented in Figure 38.

The display formula for these data is

```
dryMatter ~ fertilizer | Tray
```

based on the columns named:

`dryMatter`: dry matter amount (%) 30 days after planting.

`fertilizer`: fertilizer amount (mg).

`Tray`: a factor giving a unique identifier for each tray.

`Moisture`: moisture level (ml).

34 Wheat2 — Wheat yield trials

Stroup and Baenziger (1994) report data on an agronomic yield trial to compare 56 different varieties of wheat. The experimental units were organized according to a randomized complete block design with four blocks. All 56 varieties of wheat were used in each block. The latitude and longitude of each experimental unit in the trial were also recorded. The data, shown in Figure 39, are also analyzed in Littell et al. (1996, §9.6.2).

Columns The display formula for these data is

```
yield ~ variety | Block
```

based on the columns named:

`yield`: wheat yield.

`variety`: a factor giving the unique identifier for each wheat variety.

`Block`: a factor giving a unique identifier for each block in the experiment.

`latitude`: latitude of the experimental unit.

`longitude`: longitude of the experimental unit.

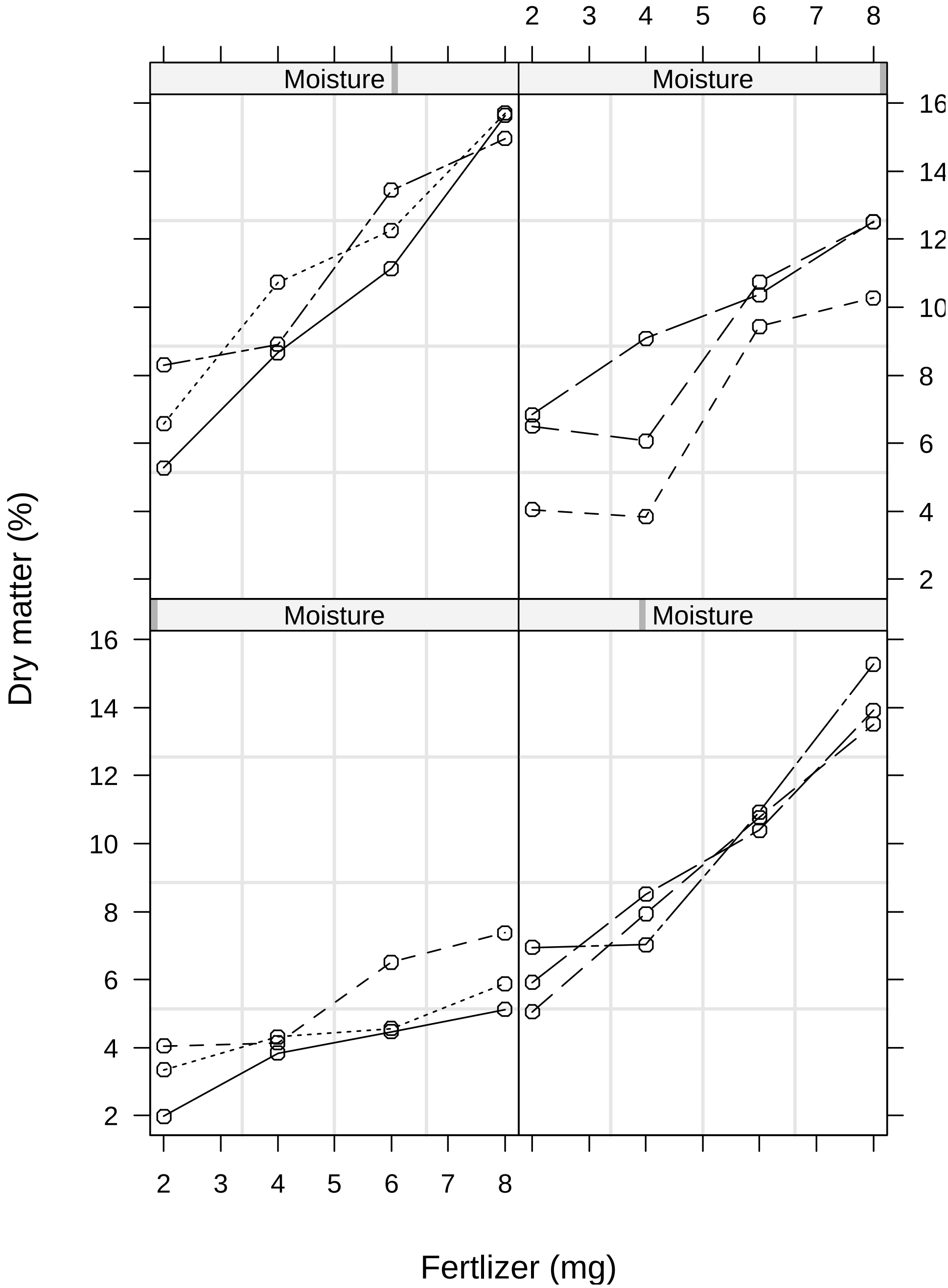

Figure 38: *Dry matter amount of wheat plants 30 days after planting, for different combinations of moisture level and fertilizer amount.*

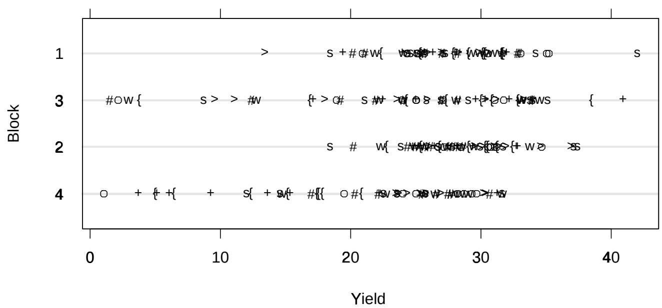

Figure 39: *Yields of 56 different varieties of wheat for each block of a randomized complete block design.*

References

Atkinson, A. C. (1985). *Plots, Transformations, and Regression*, Oxford Press, New York.

Baumann, F. and Waldvogel, F. (1963). La restitution pastsystolique de la contraction de l'oreillette gauche du rat. Effets de divers ions et de l'acetylcholine, *Helvetica Physiologica Acta* **21**.

Boeckmann, A. J., Sheiner, L. B. and Beal, S. L. (1994). *NONMEM Users Guide – Part V*, NONMEM Project Group, University of California, San Francisco.

Brillinger, D. (1987). Comment on a paper by C.R. Rao, *Statistical Science* **2**: 448–450.

Chambers, J. M. and Hastie, T. J. (eds) (1992). *Statistical Models in S*, Chapman & Hall, New York. Formerly Monterey: Wadsworth and Brooks/Cole.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*, 2nd edn, Wiley, New York.

Crowder, M. and Hand, D. (1990). *Analysis of Repeated Measures*, Chapman and Hall.

Daniel, C. and Wood, F. S. (1980). *Fitting Equations to Data*, 2nd edn, Wiley, New York.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*, 1st edn, Chapman & Hall, London.

Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences*, 4th edn, Wadsworth, Belmont, CA.

Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Oxford University Press, New York.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, 2nd edn, Wiley, New York.

Fei, Gross, Lofgren, Mora-Worms and Chen (1990). Cyclic amp response to recombinant human relaxin by cultured endometrial cells – a specific and high througput in-vitro bioassay, *Biochemical and Biophysical Research Communications* **170**: 214–222.

Grasela and Donn (1985). Neonatal population pharmacokinetics of phenobarbital derived from routine clinical data, *Developmental Pharmacology and Therapeutics* **8**: 374–0383.

Hader, R. J. and Grandage, A. H. E. (1958). Simple and multiple regression analysis, *in* V. Chew (ed.), *Experimental Designs in Industry*, Wiley, New York, pp. 108–137.

Hand, D. and Crowder, M. (1996). *Practical Longitudinal Data Analysis*, Texts in Statistical Science, Chapman and Hall.

Hicks, C. R. (1993). *Fundamental Concepts in the Design of Experiments*, 4th edn, Harcourt Brace & Company, New York.

Joyner and Boore (1981). Peak horizonal acceleration and velocity from strong-motion records including records from the 1979 imperial valley, california, earthquake, *Bulletin of the Seismological Society of America* **71**: 2011–2038.

Kung, F. H. (1986). Fitting logistic growth curve with predetermined carrying capacity, *ASA Proceedings of the Statistical Computing Section* pp. 340–343.

Kwan, Breault, Umbenhauer, McMahon and Duggan (1976). Kinetics of indomethicin absorption, elimination, and enterohepatic circulation in man, *Journal of Pharmacokinetics and Biopharmaceutics* **4**: 255–280.

Linder, A., Chakravarti, I. M. and Vuagnat, P. (1964). Fitting asymptotic regression curves with different asymptotes, *in* C. R. Rao (ed.), *Contributions to Statistics*, Pergamon Press, Oxford, pp. 221–228.

Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. (1996). *SAS System for Mixed Models*, SAS Institute Inc., Cary, NC.

Lu and Meeker (1993). Using degradation measures to estimate a time-to-failure distribution, *Technometrics* **35**(2): 161–174.

Ludbrook, J. (1994). Repeated measurements and multiple comparisons in cardiovascular research, *Cardiovascular Research* **28**: 303–311.

Milliken, G. A. and Johnson, D. E. (1992). *Analysis of messy data. Volume 1: Designed experiments*, Chapman & Hall.

Paul, P. C. (1943). PhD thesis, Iowa State College.

Pierson and Ginther (1987). Follicular population dynamics during the estrus cycle of the mare, *Animal Reproduction Science* **14**: 219–231.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**: 313–326.

Potvin, C., Lechowicz, M. J. and Tardif, S. (1990). The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures, *Ecology* **71**: 1389–1400.

Prater, N. H. (1955). Estimate gasoline yields from crudes, *Petroleum Refiner* **35**(5).

Stroup, W. W. and Baenziger, P. S. (1994). Removing spatial variation from wheat yield trials: a comparison of methods, *Crop Science* **34**: 62–66.

Venables, W. N. and Ripley, B. D. (1996). Complements to Modern Applied Statistics with S-Plus. Supplement to the current edition of MASS covering new features.

Verme, Ludden, Celmenti and Harris (1992). Pharmacokinetics of quinidine in male patients: A population analysis, *Clinical Pharmacokinetics* **22**: 468–480.

Vonesh, E. F. and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures, *Biometrics* **48**: 1–18.

Yates, F. (1935). Complex experiments, *Journal of the Royal Statistical Society (Supplement)* **2**: 181–247.