Chapter 1

Motivating Data Sets

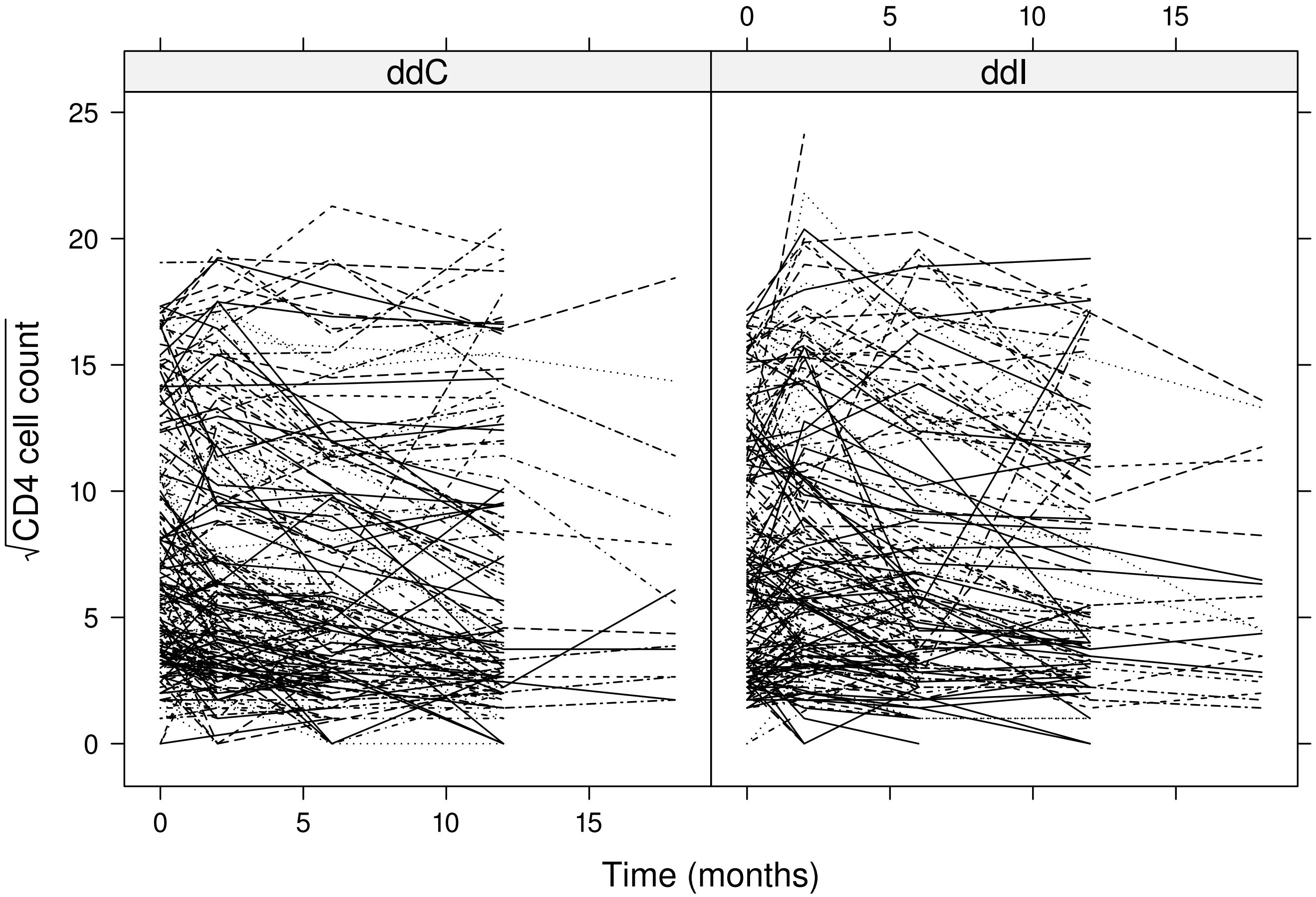
1.1 Motivating Longitudinal Studies

- **AIDS:** 467 HIV infected patients who had failed or were intolerant to zidovudine therapy (AZT) (Abrams et al., NEJM, 1994)

- The aim of this study was to compare the efficacy and safety of two alternative antiretroviral drugs, didanosine (ddI) and zalcitabine (ddC)

- Outcomes of interest:
  - ▷ CD4 cell count measurements at baseline, 2, 6, 12 and 18 months
  - ▷ randomized treatment: 230 patients ddI and 237 ddC
  - ▷ prevOI: previous opportunistic infections

1.1 Motivating Longitudinal Studies (cont'd)



1.1 Motivating Longitudinal Studies (cont'd)

- $\boxed{\text{Research Questions:}}$
  - ▷ How CD4 cell count evolves in time for this cohort of patients?
  - ▷ Does treatment improve average longitudinal evolutions?

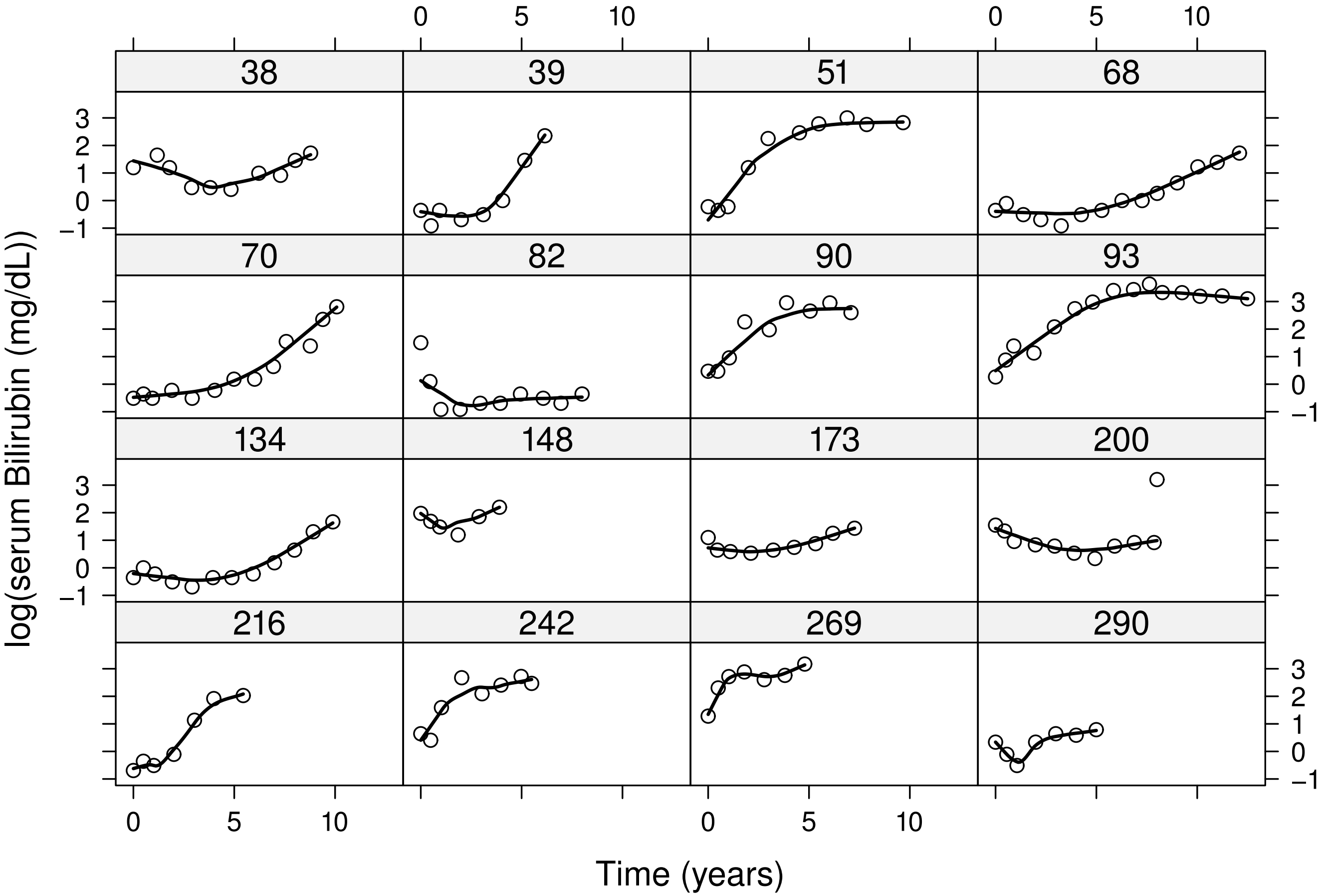
1.1 Motivating Longitudinal Studies (cont'd)

- **PBC:** Primary Biliary Cirrhosis:
  - ▷ a chronic, fatal but rare liver disease
  - ▷ characterized by inflammatory destruction of the small bile ducts within the liver

- Data collected by Mayo Clinic from 1974 to 1984 (Murtaugh et al., Hepatology, 1994)

- Outcomes of interest:
  - ▷ longitudinal serum bilirubin, serum cholesterol, prothrombin time
  - ▷ randomized treatment: 158 patients received D-penicillamine and 154 placebo

1.1 Motivating Longitudinal Studies (cont'd)



1.1 Motivating Longitudinal Studies (cont'd)

- $\boxed{\text{Research Questions:}}$
  - ▷ Do men have higher serum bilirubin during follow-up than women?
  - ▷ Is there a difference in the average longitudinal evolutions of serum bilirubin when we correct for age differences at baseline and gender differences during follow-up?

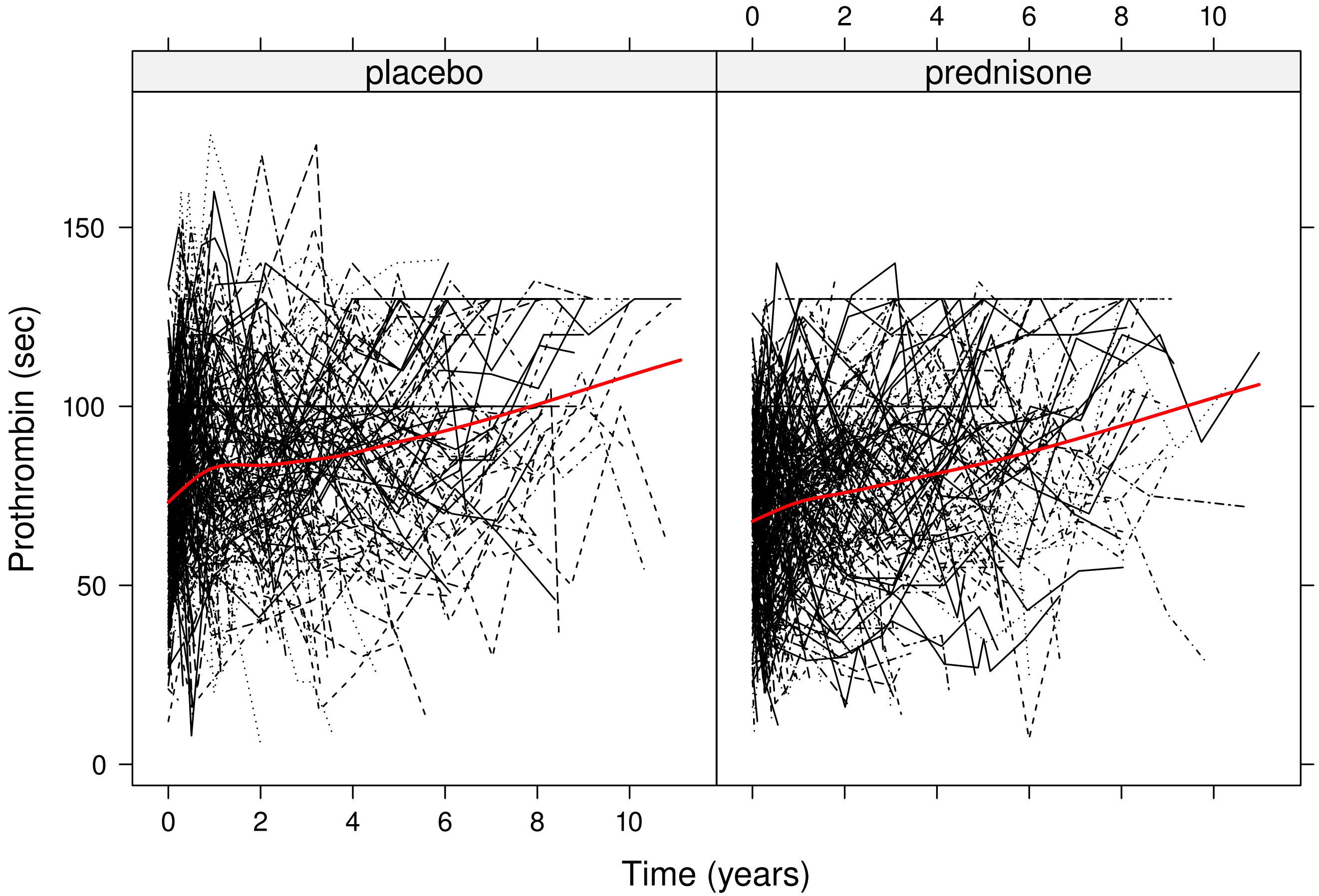
1.1 Motivating Longitudinal Studies (cont'd)

- **Prothro:** Prednisone versus placebo in liver cirrhosis patients
  - ▷ slowly progressing disease in which healthy liver tissue is replaced with scar tissue, eventually preventing the liver from functioning properly

- Randomized trial in Denmark (Andersen et al., Springer, 1993)

- Outcomes of interest:
  - ▷ randomized treatment: 158 patients received D-penicillamine and 154 placebo
  - ▷ longitudinal prothrombin times

1.1 Motivating Longitudinal Studies (cont'd)



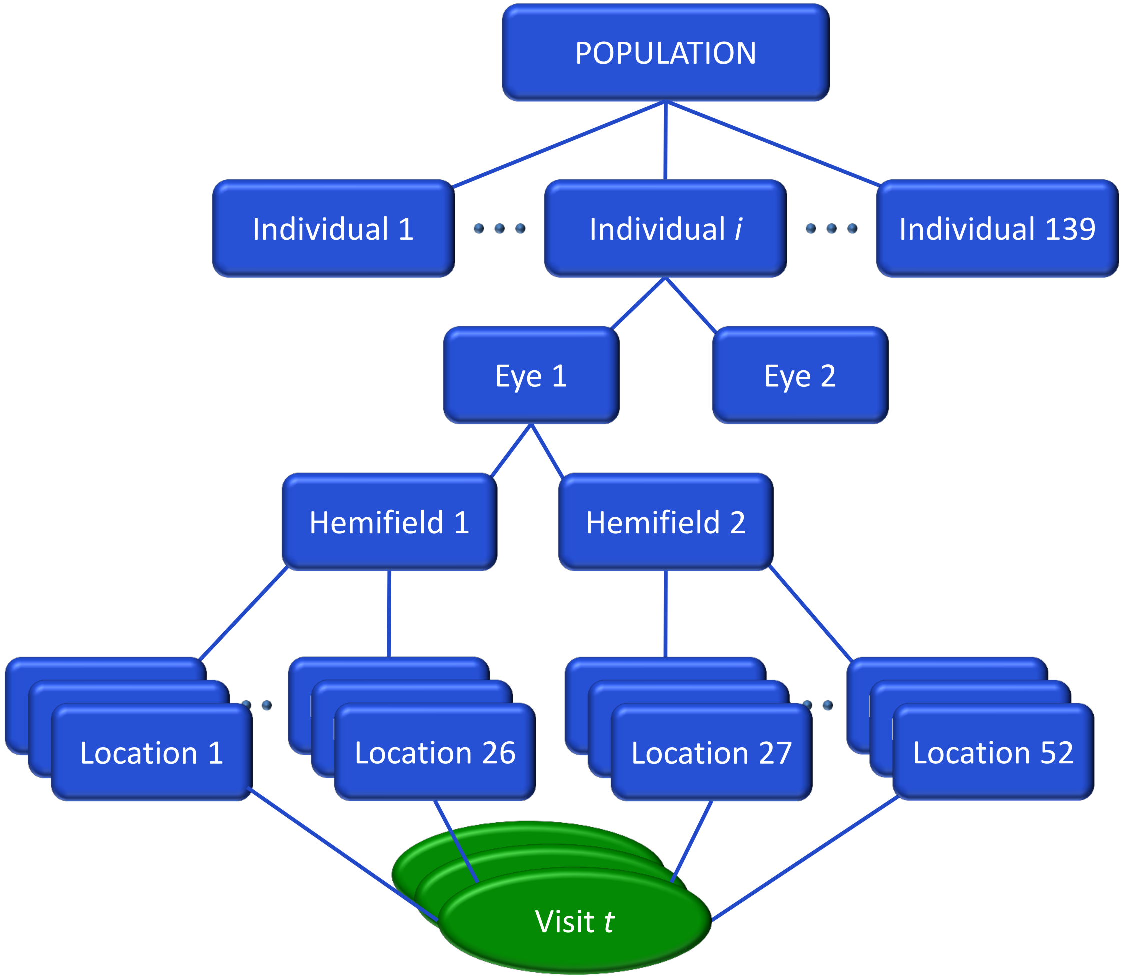
1.1 Motivating Longitudinal Studies (cont'd)

- **Glaucoma:** A group of eye conditions resulting in optic nerve damage, which may cause loss of vision

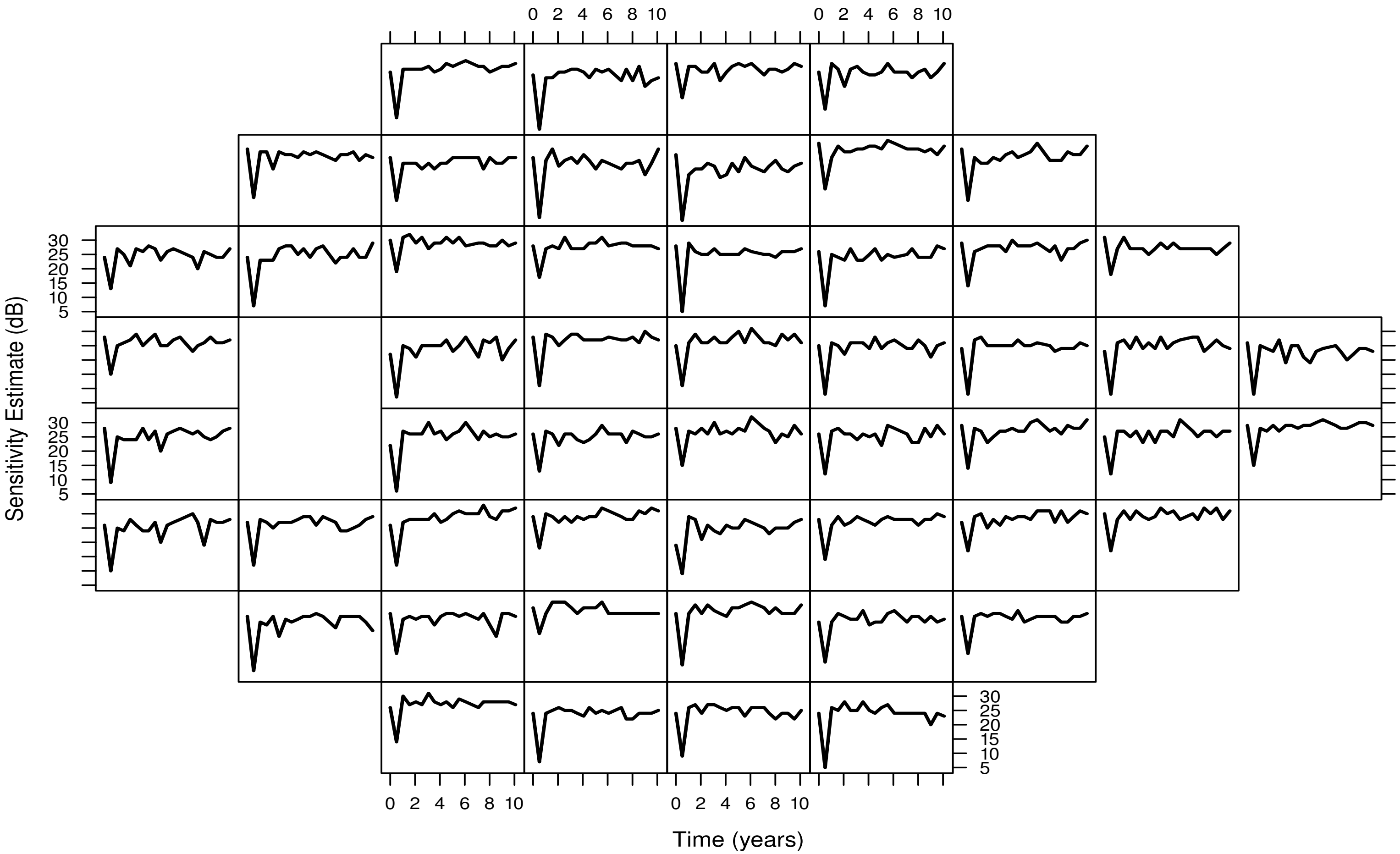
- Ongoing prospective cohort study on 139 patients (80% men) conducted by the Rotterdam Eye Hospital in the Netherlands `http://rod-rep.com`

- Outcome of interest:
  - ▷ Visual field (VF) sensitivity collected at approximately 6-months intervals

1.1 Motivating Longitudinal Studies (cont'd)



1.1 Motivating Longitudinal Studies (cont'd)



1.1 Motivating Longitudinal Studies (cont'd)

- Research Questions:
    - ▷ Study disease progression using VF sensitivity
    - ▷ Predict rate of progression for future patients

1.2 Features of Longitudinal Data

- Repeated evaluations of the same outcome in each subject in time
  - ▷ CD4 cell count in HIV-infected patients
  - ▷ serum bilirubin in PBC patients

- Visiting process
  - ▷ some times fixed by design (e.g., in randomized trials) but often not everybody adheres to them
  - ▷ completely determined by the physicians and/or the patients

1.2 Features of Longitudinal Data (cont'd)

Measurements on the same subject are expected to be (positively) correlated

- This implies that standard statistical tools, such as the t-test and simple linear regression that assume independent observations, are not optimal for longitudinal data analysis

1.2 Features of Longitudinal Data (cont'd)

- Let's see why: The simplest case of longitudinal data are paired data

- $\boxed{\text{Example:}}$ We consider the baseline and 6-month longitudinal measurements of square root CD4 cell count from the AIDS dataset

	n	mean	sd
$month = 0$	294	7.73	4.69
$month = 6$	294	6.71	4.96

1.2 Features of Longitudinal Data (cont'd)

- There is an average decrease of about 1 unit

- The classical analysis of paired data is based on comparisons within subjects:

$$\Delta_i \;=\; Y_i(t=0) - Y_i(t=6), \qquad i = 1, \ldots, n$$

- A positive Δ_i corresponds to a decrease of the square root CD4 cell count, while a negative Δ_i is equivalent to an increase

- Testing for a time effect is now equivalent to testing whether the average difference μ_Δ equals zero

1.2 Features of Longitudinal Data (cont'd)

- The paired t-test yields

```
        Paired t-test

data:  CD4 by obstime t = 6.472, df = 293, p-value = 4.057e-10
alternative hypothesis: true difference in means is not equal to 0 95 percent
confidence interval:
 0.7105585 1.3315439
sample estimates: mean of the differences
               1.021051
```

1.2 Features of Longitudinal Data (cont'd)

- What if we had ignored the paired nature of the data?

- We then could have used a two-sample (unpaired) t-test to compare the average CD cell count at the two time points

```
       Welch Two Sample t-test

data:  CD4 by obstime t = 2.565, df = 584.229, p-value = 0.01056
alternative hypothesis: true difference in means is not equal to 0 95 percent
confidence interval:
 0.2392406 1.8028617
sample estimates: mean in group 0 mean in group 6
       7.730128        6.709077
```

1.2 Features of Longitudinal Data (cont'd)

- We would still have found a significant difference ($p = 0.0106$), but the p-value would have been many times larger compared to the one obtained using the paired t-test

- The two-sample t-test does not take into account the fact that the measurements are not independent observations

- This illustrates that classical statistical models which assume independent observations will not be optimal for the analysis of longitudinal data

1.2 Features of Longitudinal Data (cont'd)

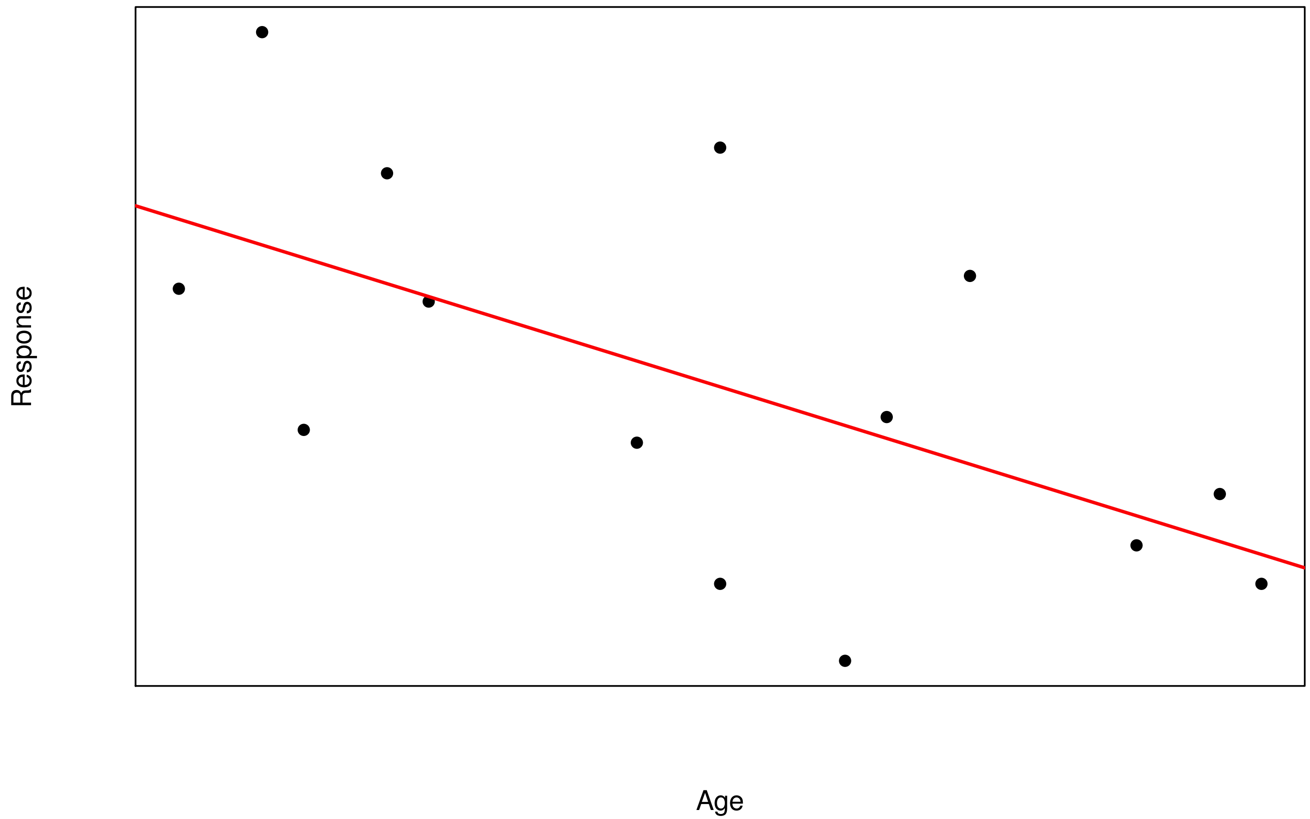
- Longitudinal studies allow to investigate

  1. how treatment means differ at specific time points, e.g., at the end of the study (*cross-sectional effect*)
  2. how treatment means or differences between means of treatments change over time (*longitudinal effect*)

- An example: Suppose it is of interest to study the relation between some response Y and age

  ▷ a cross-sectional study yields the following data:

1.2 Features of Longitudinal Data (cont'd)

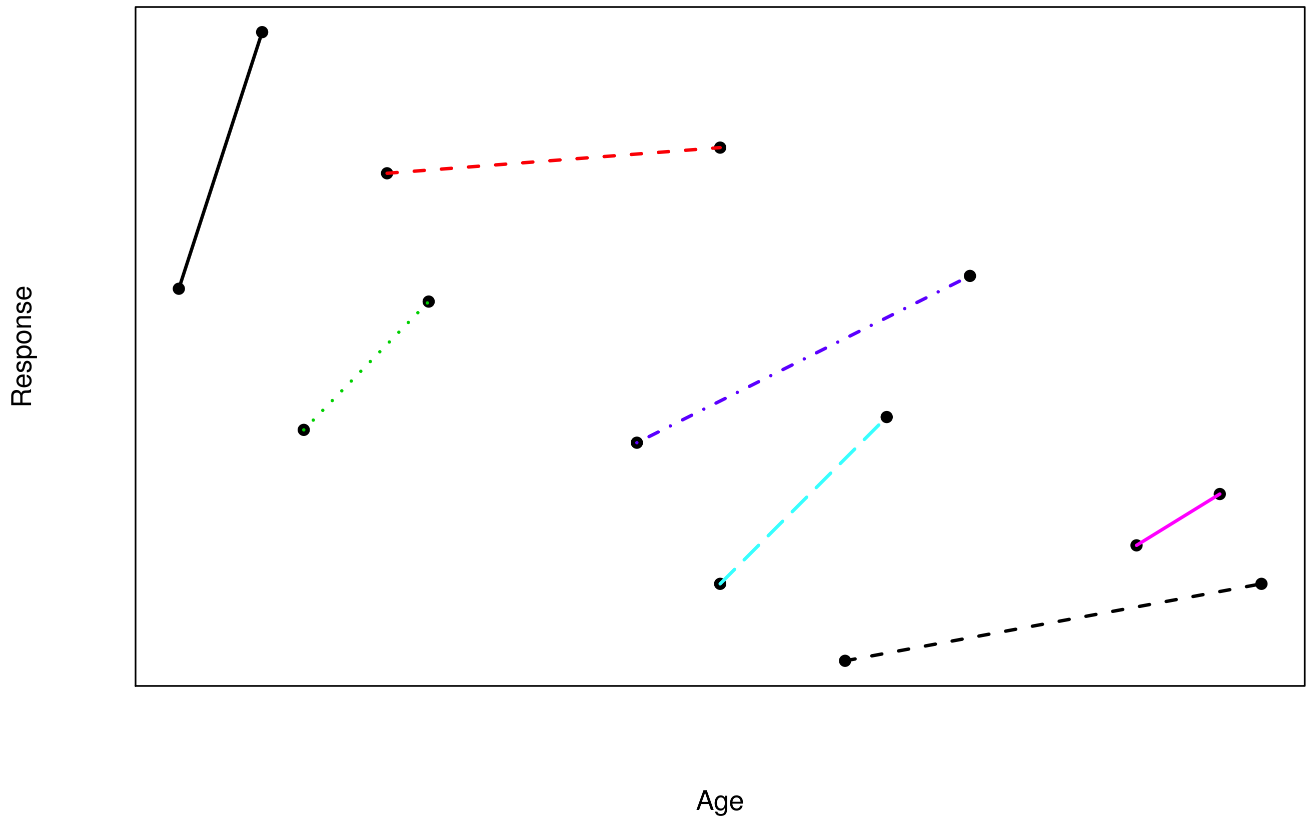


1.2 Features of Longitudinal Data (cont'd)

- The graph clearly suggests a negative relation between Y and age

- **Nevertheless**, exactly the same observations also could have been obtained in a longitudinal study, with 2 measurements per subject

1.2 Features of Longitudinal Data (cont'd)



1.2 Features of Longitudinal Data (cont'd)

Are we now still inclined to conclude that there is a negative relation between Y and age?

- Conclusion: Longitudinal data allow to distinguish differences between subjects from changes within subjects

1.3 Review of Key Points

- Grouped & longitudinal data: Features
  - ▷ measurements on the same subject are correlated
  - ▷ allow to distinguish within and between subjects effects