

Aufgabe 2

Betrachten Sie den Datensatz *Soybean* aus dem R-Paket *nlme*.

- (a) Lesen Sie die Hilfe zum Datensatz und wenden Sie die Funktionen `str()` und `plot()` auf *Soybean* an.
- (b) Nutzen Sie die Funktion `xypplot()` aus dem R-Paket *lattice*, um eine Graphik des logarithmierten Gewichtes in Abhängigkeit von der Sorte (*Variety*) und der Parzelle (*Plot*) zu erstellen. Warum sollte man das logarithmierte Gewicht glatt und nicht linear modellieren?
- (c) Fitten Sie ein Modell für das logarithmierte Gewicht mit Jahr, Sorte und glatter Funktion der Zeit (gleiche Form für alle Plots), sowie Plot-spezifischen zufälligen Intercepts. Lassen Sie sich auch eine geeignete Modellzusammenfassung ausgeben.
Hinweis: Verwenden Sie die Funktion `gamm()` aus dem R-Paket *mgcv*.
- (d) Fügen Sie dem Modell aus (c) eine zufällige Steigung pro Plot hinzu.
- (e) Schätzen Sie ein Modell für das logarithmierte Gewicht mit glatter Funktion der Zeit für jede Sorte, sowie einem festen Effekt für Jahr und zufälligen Intercepts für die Plots.

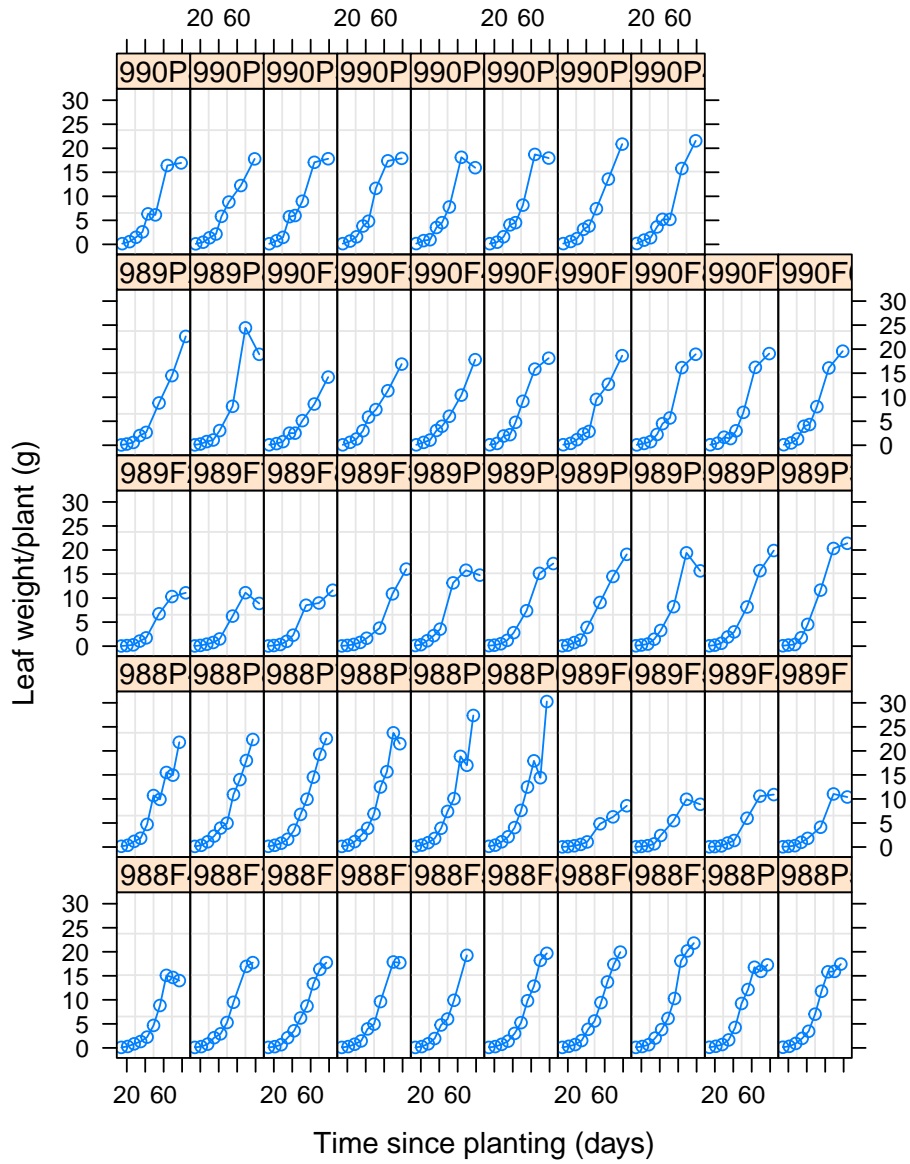
Lösung 2

```
(a) library(mgcv)
    library(lattice)
    library(nlme)

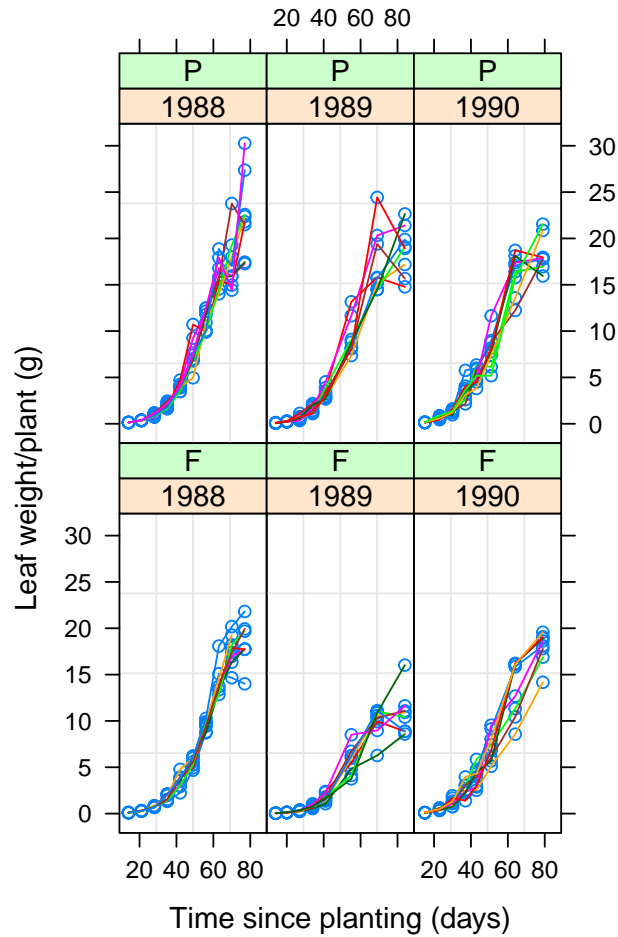
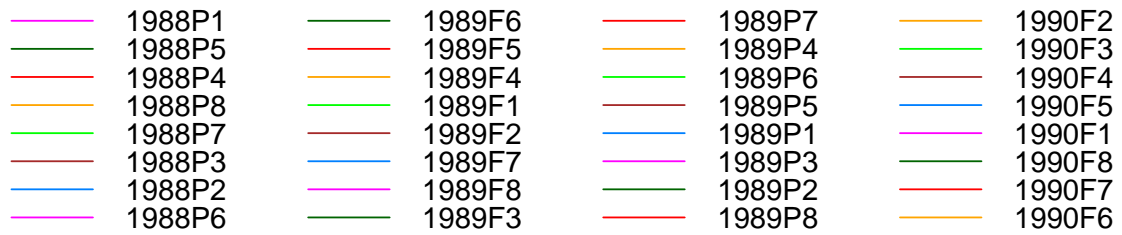
    # ?Soybean
    # str(Soybean, 1)
    class(Soybean)

    [1] "nfnGroupedData" "nfGroupedData" "groupedData"    "data.frame"

    plot(Soybean)
```

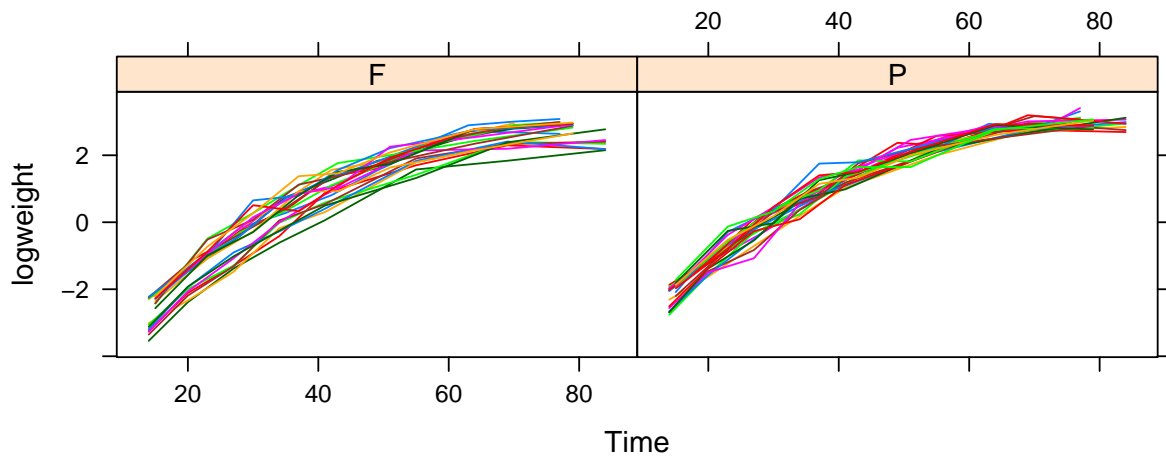


```
plot(Soybean, outer= ~ Year*Variety)
```



(b) Spaghetti-Plot pro Variety

```
Soybean$logweight <- log(Soybean$weight)
xyplot(logweight ~ Time | Variety, Soybean, groups = Plot, type = "l")
```



(c) Modelfit mit gamm

```
g1 <- gamm(logweight ~ Year + Variety +
           s(Time, bs = "ps", m = c(2,1), k = 20),
           random = list(Plot = ~ 1), data = Soybean, method = "REML")
str(g1,1)
```

List of 2

\$ lme:List of 18

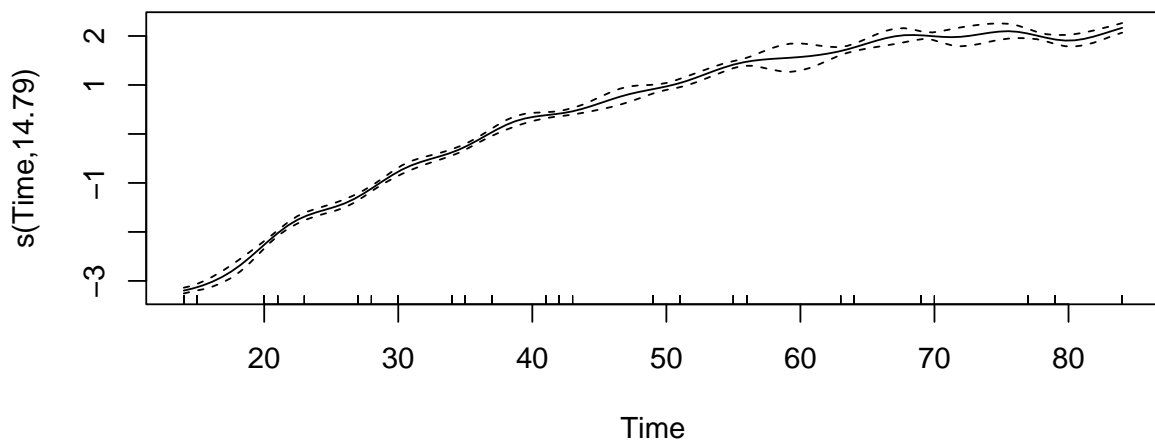
..- attr(*, "class")= chr "lme"

\$ gam:List of 31

..- attr(*, "class")= chr "gam"

- attr(*, "class")= chr [1:2] "gamm" "list"

```
plot(g1$gam)
```



```

# summary(g1flme)
# summary(g1fgam)
coef(g1$gam)

(Intercept)   Year1989   Year1990   VarietyP   s(Time).1   s(Time).2
0.75586820 -0.46108969 0.06288942 0.34509745 -2.27338379 -2.05151120
s(Time).3   s(Time).4   s(Time).5   s(Time).6   s(Time).7   s(Time).8
-0.68439390 -0.64728818 0.29171744 0.50666203 1.27157065 1.30077059
s(Time).9   s(Time).10  s(Time).11  s(Time).12  s(Time).13  s(Time).14
1.33624054 1.68802782 2.25642026 1.90023067 2.28915763 2.79972151
s(Time).15  s(Time).16  s(Time).17  s(Time).18  s(Time).19
2.36257107 2.69208863 2.30023472 2.60262022 2.45607553

```

Alternativ über gam:

```

g1gam <- gam(logweight ~ Year + Variety +
             s(Time, bs = "ps", m = c(2,1), k = 20) +
             s(Plot, bs = "re"), data = Soybean, method = "REML")
# summary(g1gam)
coef(g1gam)[1:23]

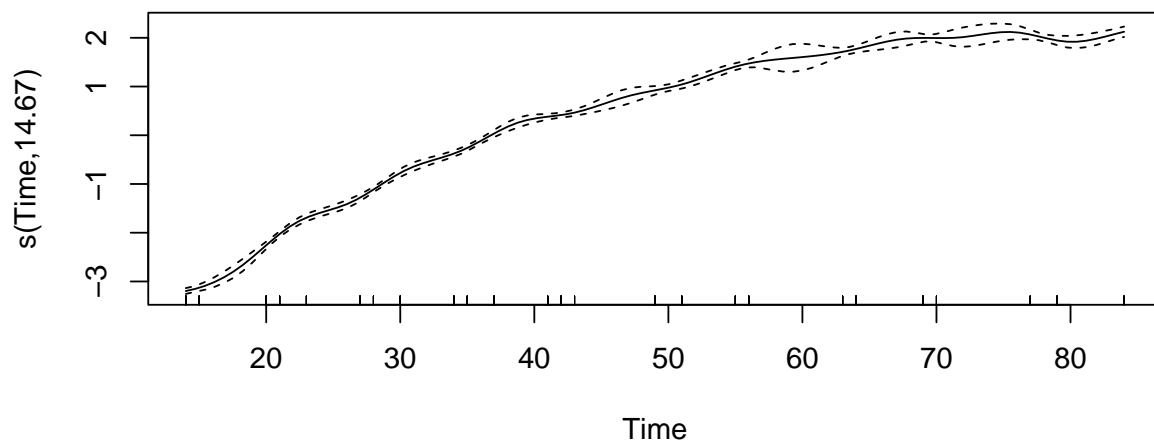
(Intercept)   Year1989   Year1990   VarietyP   s(Time).1   s(Time).2
0.75586820 -0.46108969 0.06288943 0.34509745 -2.27338383 -2.05151111
s(Time).3   s(Time).4   s(Time).5   s(Time).6   s(Time).7   s(Time).8
-0.68439399 -0.64728811 0.29171736 0.50666208 1.27157057 1.30077063
s(Time).9   s(Time).10  s(Time).11  s(Time).12  s(Time).13  s(Time).14
1.33624046 1.68802786 2.25642023 1.90023071 2.28915765 2.79972138
s(Time).15  s(Time).16  s(Time).17  s(Time).18  s(Time).19
2.36257121 2.69208852 2.30023480 2.60262020 2.45607547

```

```

(d) g2 <- gamm(logweight ~ Year + Variety +
              s(Time, bs = "ps", m = c(2,1), k = 20),
              random = list(Plot = ~ 1 + Time), data = Soybean, method = "REML")
# summary(g2)
plot(g2$gam)

```



```
coef(g2$gam)
```

(Intercept)	Year1989	Year1990	VarietyP	s(Time).1	s(Time).2
0.75472532	-0.43200992	0.06258315	0.32953429	-2.31106482	-2.04450802
s(Time).3	s(Time).4	s(Time).5	s(Time).6	s(Time).7	s(Time).8
-0.71600349	-0.65361207	0.26469721	0.49544814	1.24589490	1.28643910
s(Time).9	s(Time).10	s(Time).11	s(Time).12	s(Time).13	s(Time).14
1.33404239	1.68274186	2.21956256	1.93826004	2.30307756	2.71829139
s(Time).15	s(Time).16	s(Time).17	s(Time).18	s(Time).19	
2.40957275	2.69353255	2.31805662	2.54333720	2.38108734	

```
(e) g3 <- gamm(logweight ~ Year +
               s(Time, by = Variety, bs = "ps", m = c(2,1), k = 20),
               random = list(Plot = ~ 1), data = Soybean, method = "REML")
# summary(g3)
par(mfrow=c(1,2))
# plot(g3$gam)
# coef(g3$gam)
```