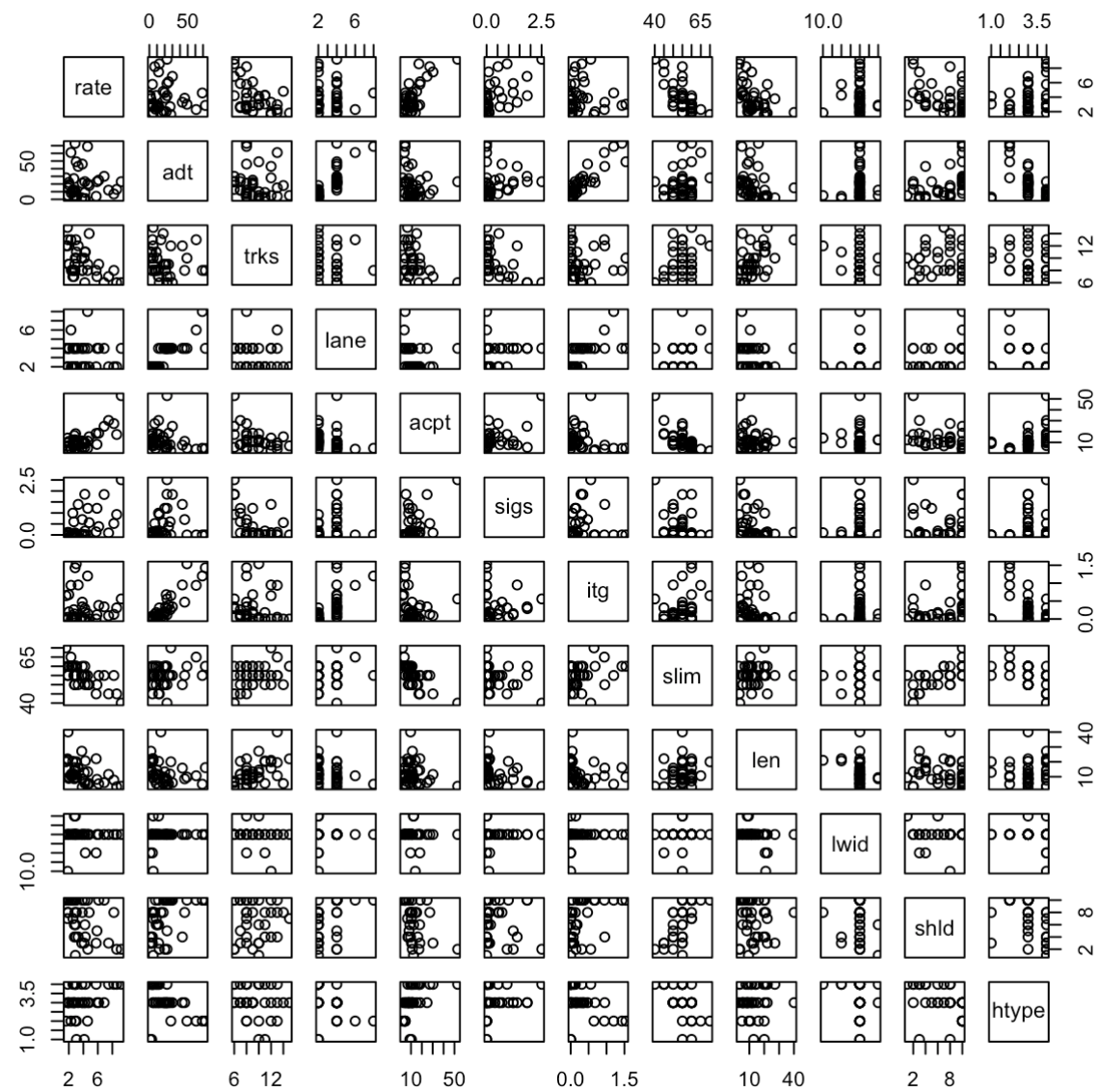


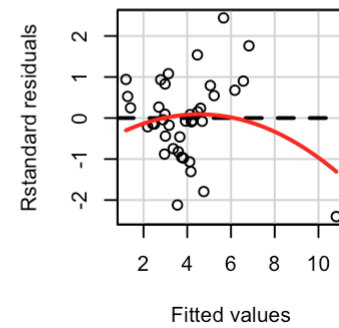
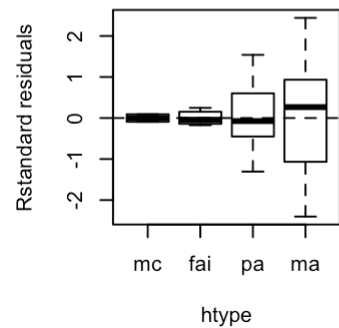
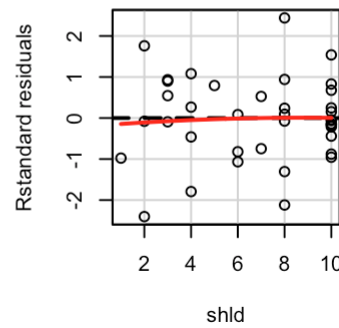
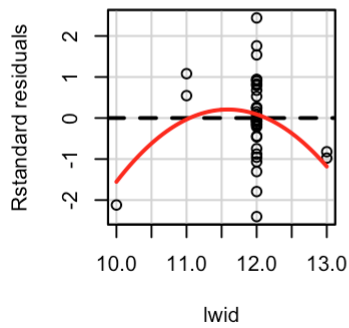
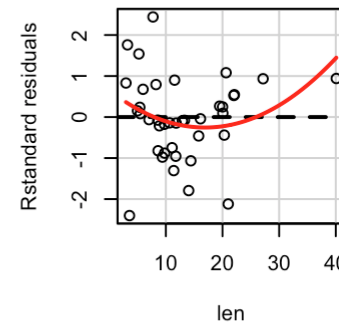
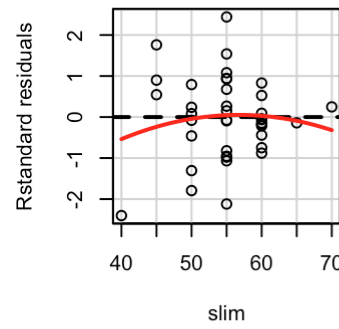
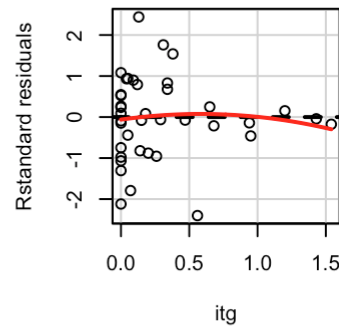
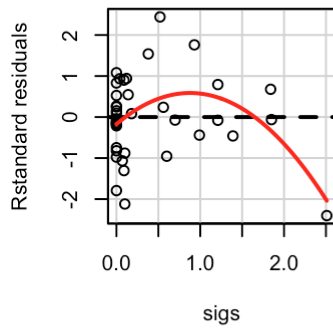
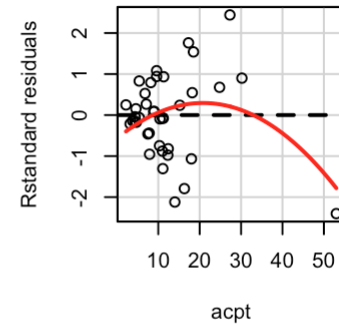
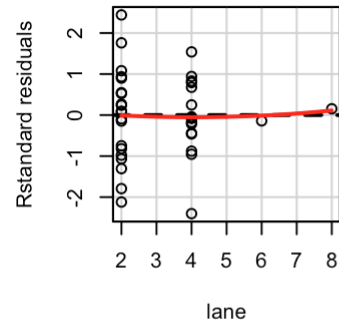
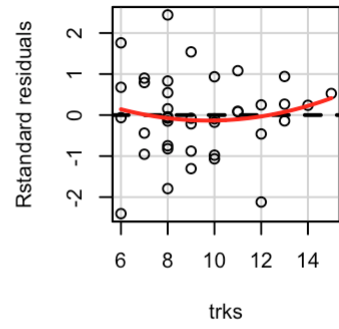
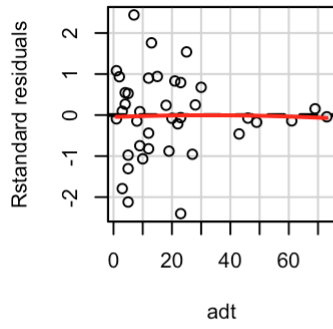
MLR: Model Selection in R

Math 430, Winter 2017

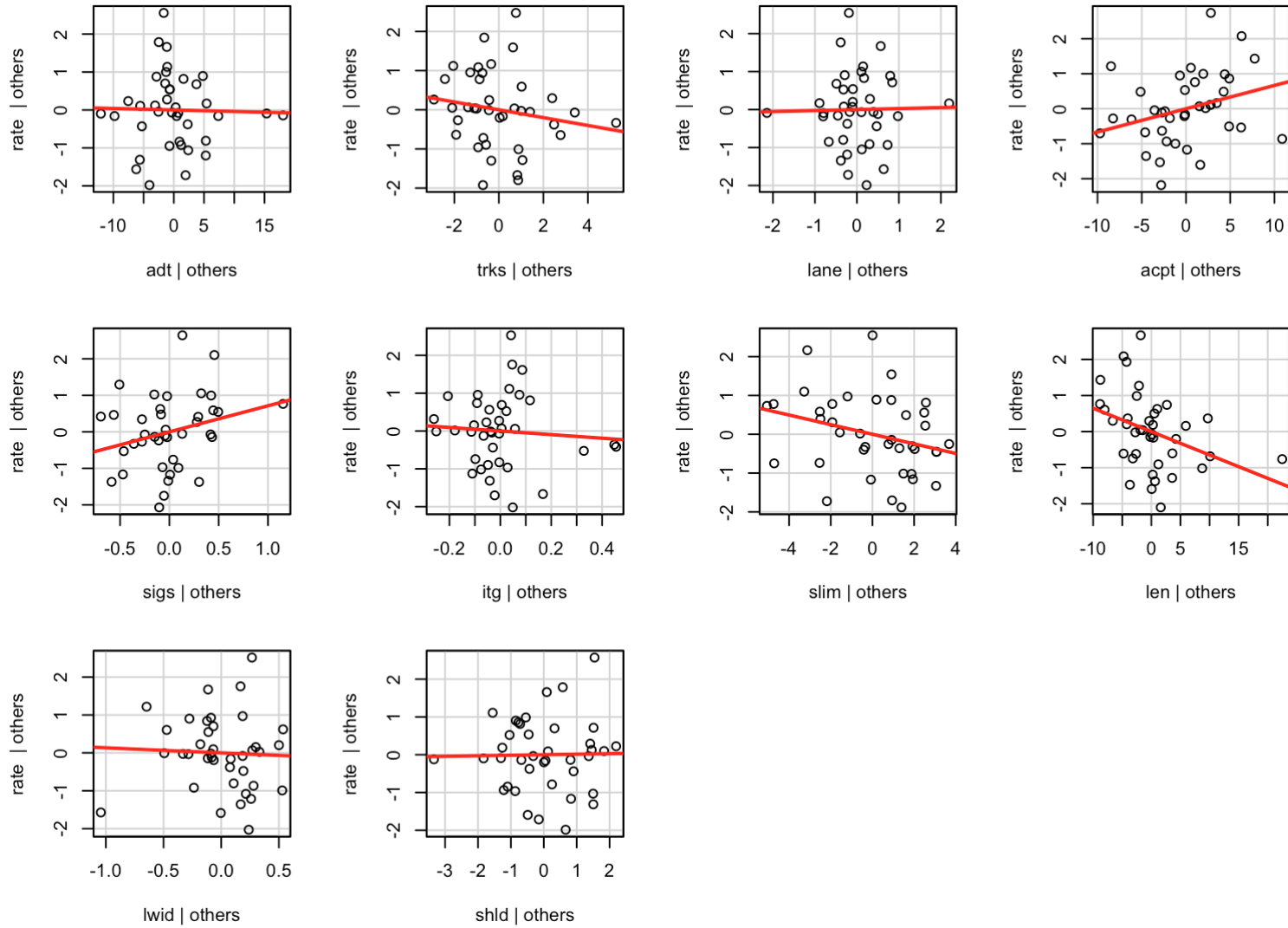
Highway accident data

Variable	Description
adt	average daily traffic count (thousands)
trks	truck volume as a percent of the total volume
lane	total number of lanes of traffic
acpt	number of access points per mile
sigs	number of signalized interchanges per mile
itg	number of freeway-type interchanges per mile
slim	speed limit
len	length of the Highway segment (miles)
lwid	lane width (feet)
shld	width in feet of outer shoulder on the roadway
htype	type of roadway/funding source
rate	accident rate per million vehicle miles

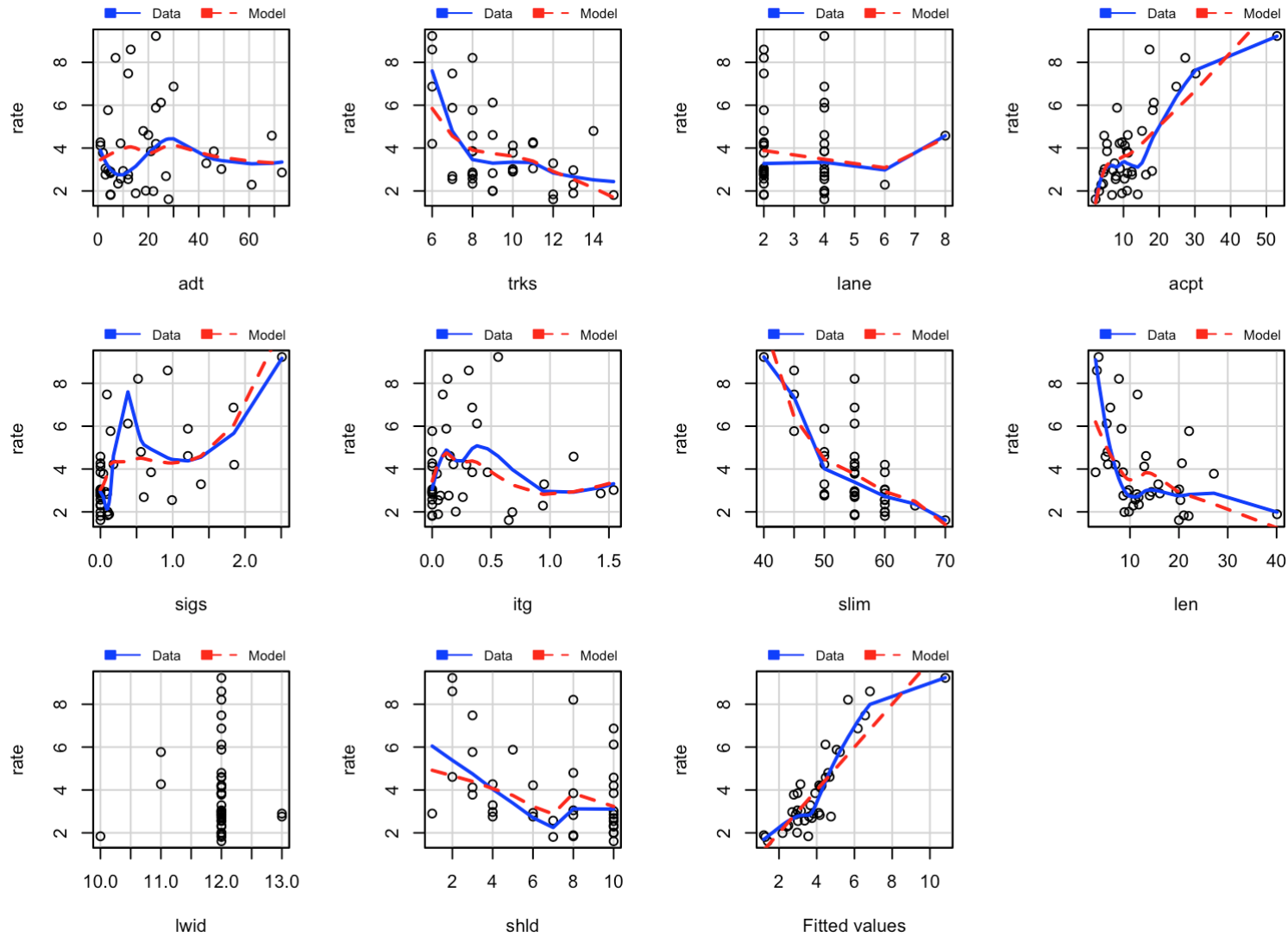


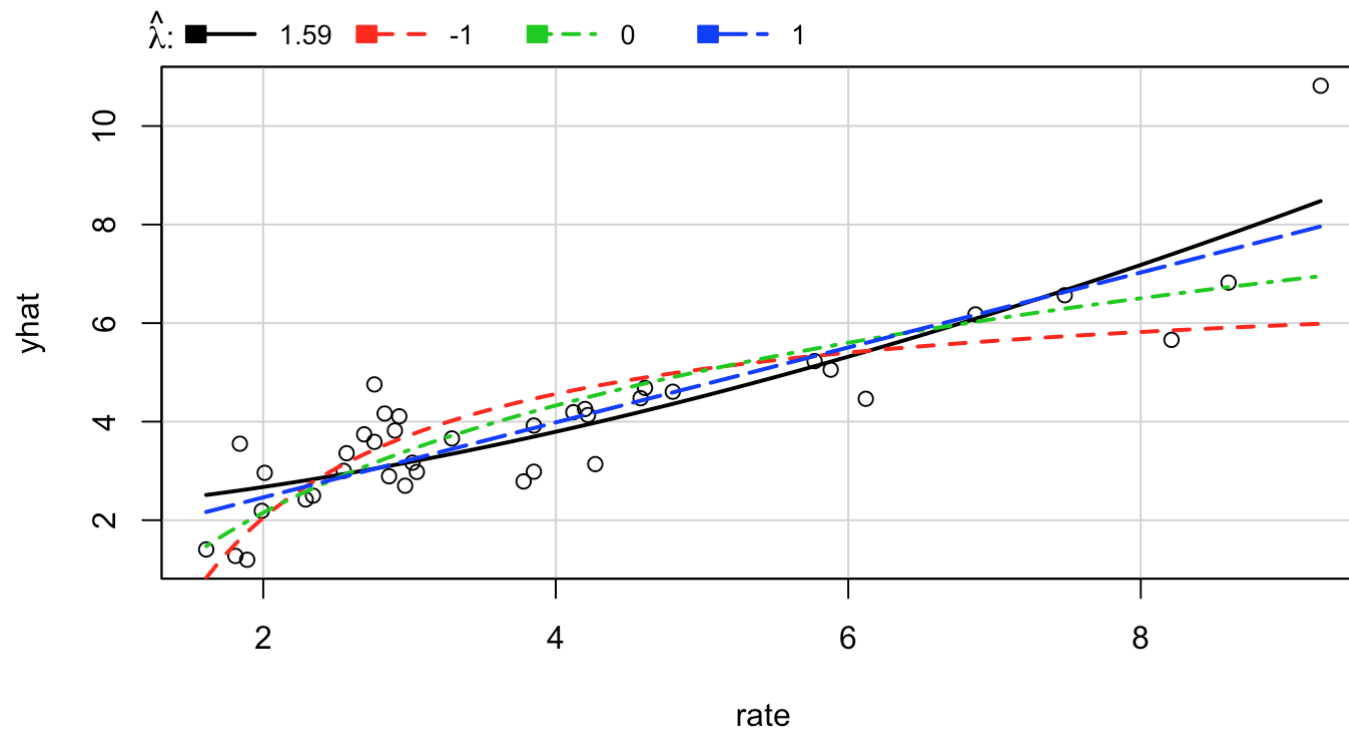


Added-Variable Plots



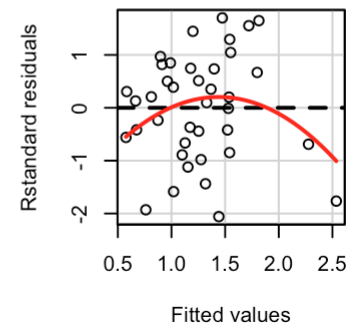
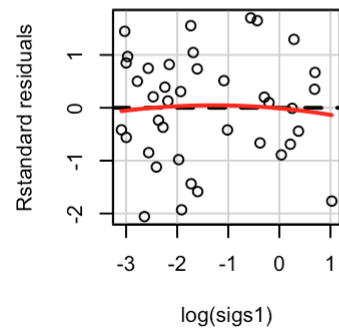
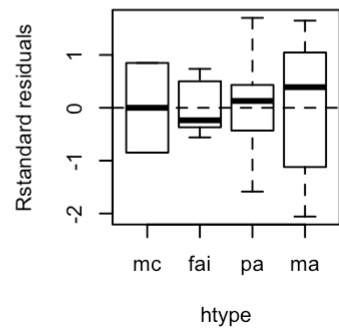
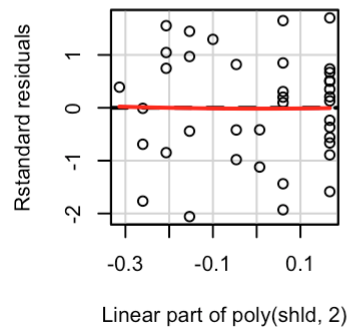
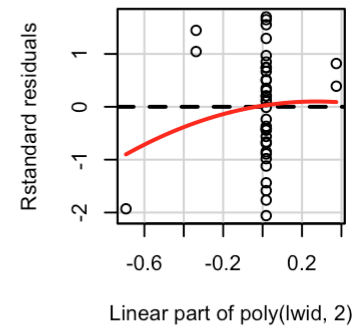
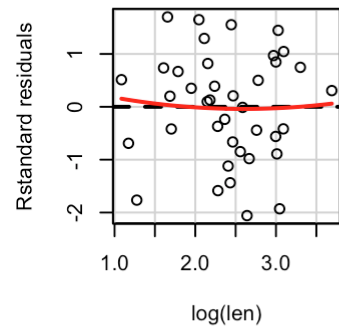
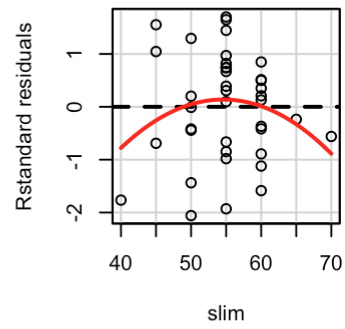
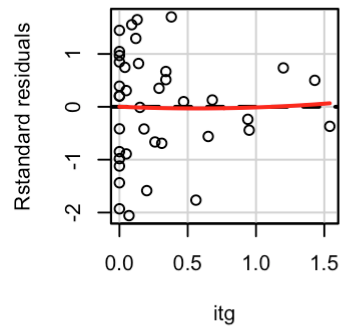
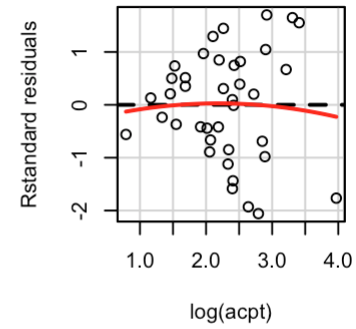
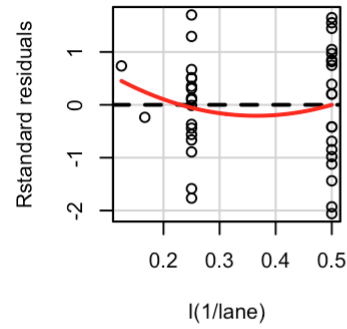
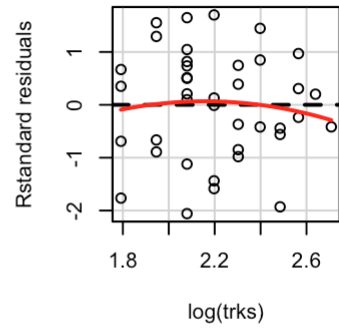
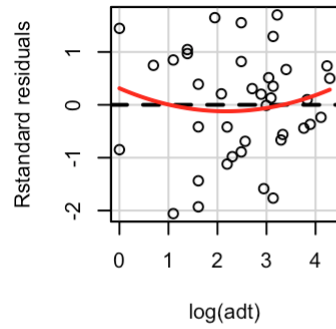
Marginal Model Plots



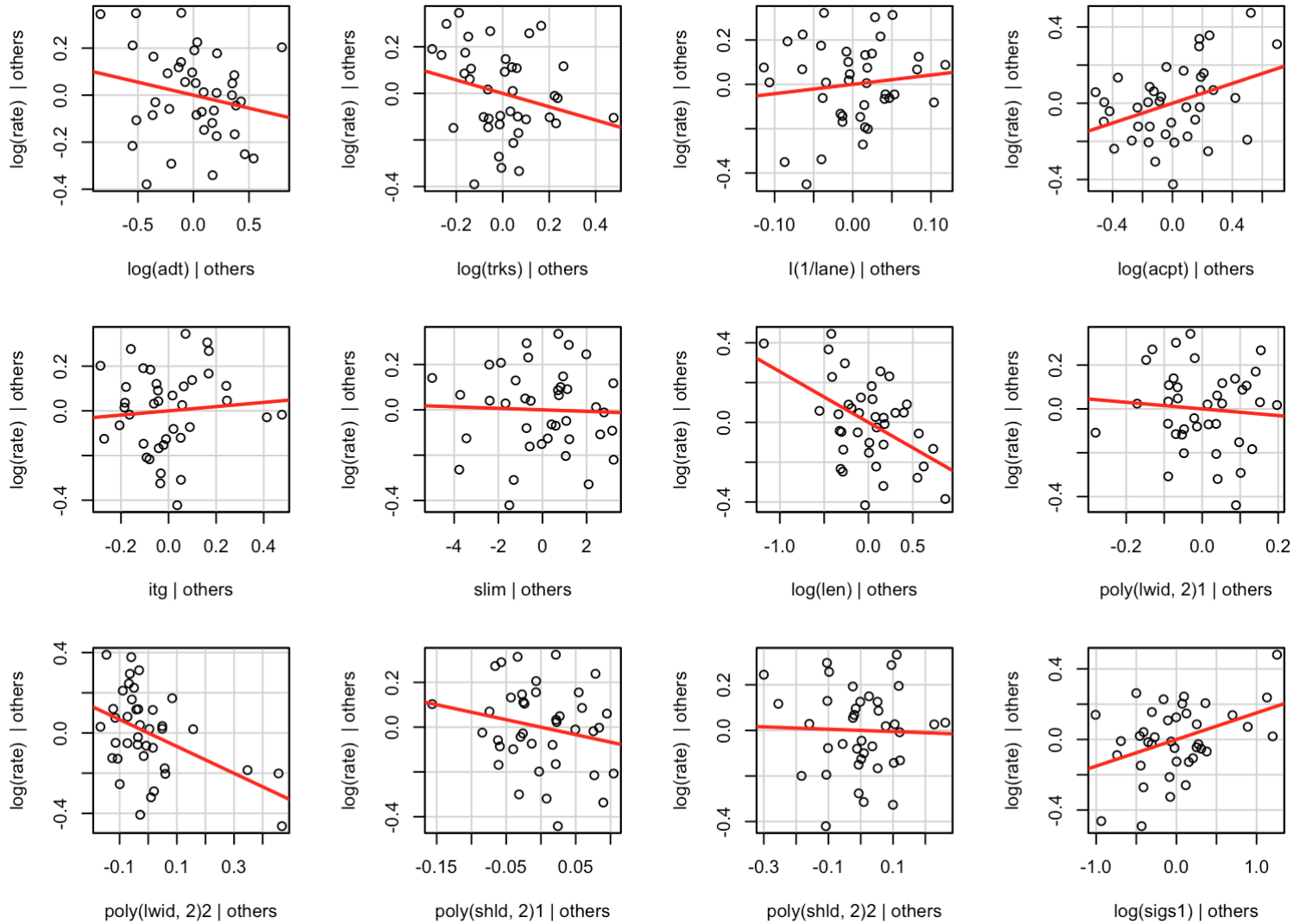


##	lambda	RSS
## 1	1.594529	26.41210
## 2	-1.000000	45.59093
## 3	0.000000	33.75636
## 4	1.000000	27.29810

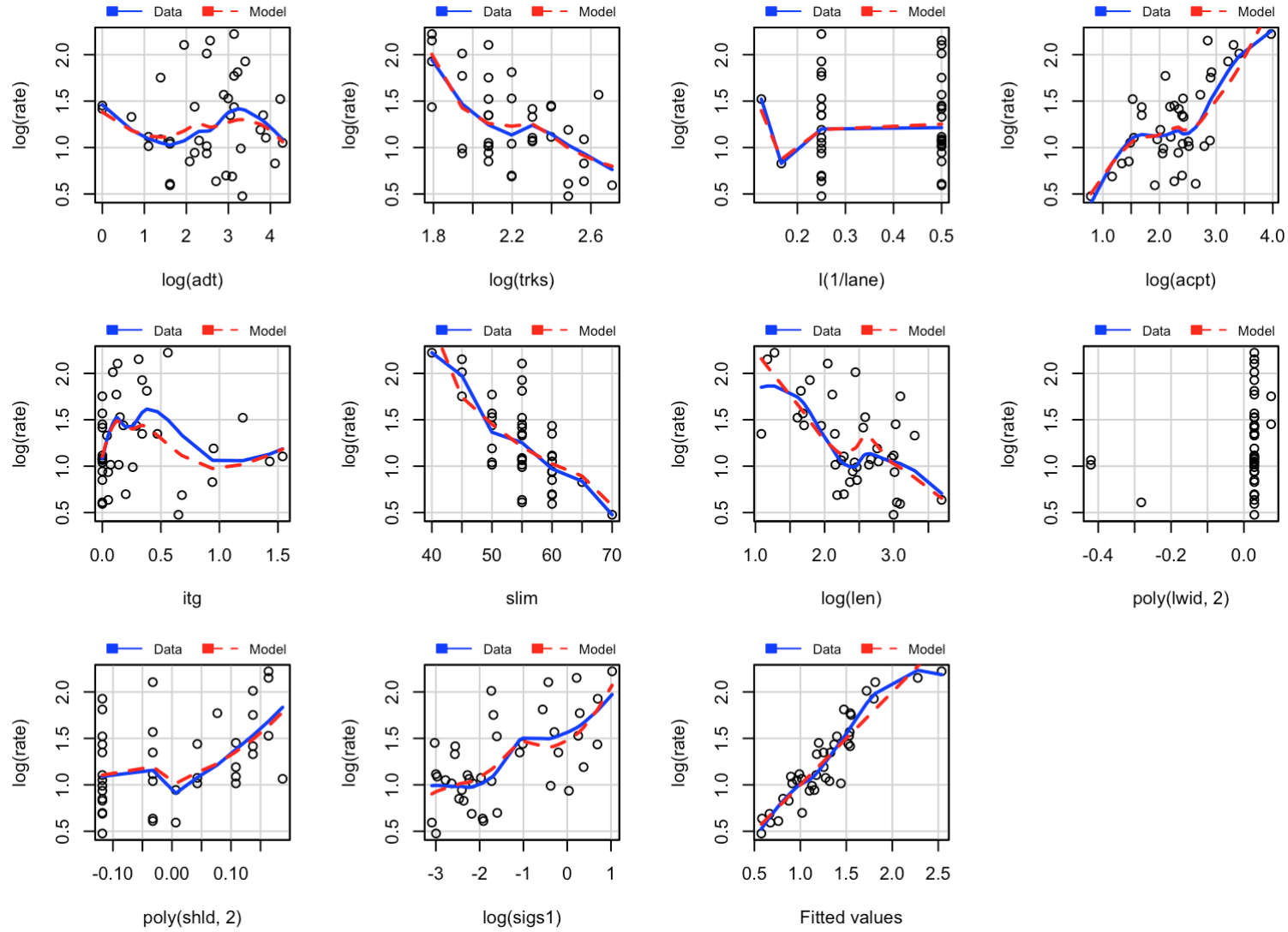

```
Highway <- mutate(Highway, sigs1 = (sigs * len + 1)/len)
full_mod_tform <- lm(log(rate) ~ log(adtl) + log(trks) + I(1/lanel) + log(acpt) +
  itg + slim + log(len) + poly(lwid, 2) + poly(shld, 2) + htype + log(sigs1),
  data = Highway)
```



Added-Variable Plots



Marginal Model Plots



The **step** command

Backward elimination

```
belim <- step(full_mod_tform, scope = list(lower = ~ 1), direction = "backward")
```

```
broom::tidy(belim)
```

##	term	estimate	std.error	statistic	p.value
## 1	(Intercept)	3.24639448	0.75119775	4.3216244	0.0001764473
## 2	log(adl)	-0.14429407	0.07746273	-1.8627547	0.0730195963
## 3	log(acpt)	0.18987179	0.10707212	1.7733075	0.0870567742
## 4	slim	-0.02011261	0.01007683	-1.9959263	0.0557516497
## 5	log(len)	-0.25644916	0.07871784	-3.2578279	0.0029403083
## 6	poly(lwid, 2)1	0.13688282	0.25106602	0.5452065	0.5899285279
## 7	poly(lwid, 2)2	-0.60177023	0.23510121	-2.5596220	0.0161662281
## 8	htypefai	0.33059140	0.33000676	1.0017716	0.3250331856
## 9	htypepa	-0.21786065	0.21955592	-0.9922786	0.3295598277
## 10	htypema	-0.06105924	0.18951707	-0.3221833	0.7497070874
## 11	log(sigs1)	0.17789568	0.05689946	3.1264916	0.0040983118

Forward selection

```
null_mod <- lm(slog(rate) ~ 1, data = Highway)
fselect <- step(null_mod, scope = list(lower = ~ 1,
upper = ~ log(adl) + log(trks) + lane + acpt + itg + slim + log(len) +
      lwid + shld + htype + sigsl),
direction = "forward")
```

```
broom::tidy(fselect)
```

##	term	estimate	std.error	statistic	p.value
## 1	(Intercept)	4.16654113	0.741064508	5.622373	2.666474e-06
## 2	slim	-0.03185201	0.010261763	-3.103951	3.832823e-03
## 3	log(len)	-0.23573454	0.084896763	-2.776720	8.867186e-03
## 4	acpt	0.01100449	0.006669289	1.650025	1.081474e-01
## 5	log(trks)	-0.32903691	0.213483661	-1.541274	1.325068e-01

Stepwise selection

```
step_hwy <- step(null_mod, scope = list(lower = ~ 1,  
upper = ~ log(adtl) + log(trks) + lane + acpt + itg + slim + log(len) +  
      lwid + shld + htype + sigsl),  
direction = "both")
```

```
broom::tidy(step_hwy)
```

##	term	estimate	std.error	statistic	p.value
## 1	(Intercept)	4.16654113	0.741064508	5.622373	2.666474e-06
## 2	slim	-0.03185201	0.010261763	-3.103951	3.832823e-03
## 3	log(len)	-0.23573454	0.084896763	-2.776720	8.867186e-03
## 4	acpt	0.01100449	0.006669289	1.650025	1.081474e-01
## 5	log(trks)	-0.32903691	0.213483661	-1.541274	1.325068e-01

Using BIC rather than AIC

```
belim_bic <- step(full_mod_tform, scope = list(lower = ~ 1), direction = "backward",  
               k = log(nrow(Highway)))
```

```
broom::tidy(belim_bic)
```

##	term	estimate	std.error	statistic	p.value
## 1	(Intercept)	3.24639448	0.75119775	4.3216244	0.0001764473
## 2	log(adl)	-0.14429407	0.07746273	-1.8627547	0.0730195963
## 3	log(acpt)	0.18987179	0.10707212	1.7733075	0.0870567742
## 4	slim	-0.02011261	0.01007683	-1.9959263	0.0557516497
## 5	log(len)	-0.25644916	0.07871784	-3.2578279	0.0029403083
## 6	poly(lwid, 2)1	0.13688282	0.25106602	0.5452065	0.5899285279
## 7	poly(lwid, 2)2	-0.60177023	0.23510121	-2.5596220	0.0161662281
## 8	htypefai	0.33059140	0.33000676	1.0017716	0.3250331856
## 9	htypepa	-0.21786065	0.21955592	-0.9922786	0.3295598277
## 10	htypema	-0.06105924	0.18951707	-0.3221833	0.7497070874
## 11	log(sigs1)	0.17789568	0.05689946	3.1264916	0.0040983118

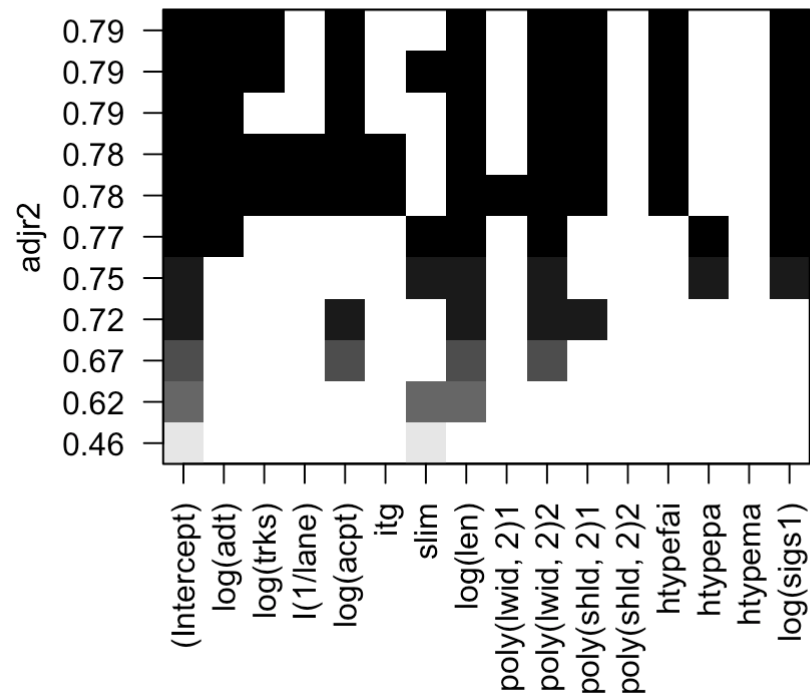
The **regsubsets** command

All subsets in R

```
library(leaps)
regfit_full <- regsubsets(log(rate) ~ log(adtt) + log(trks) + I(1/lan) + log(acpt) +
  itg + slim + log(len) + poly(lwid, 2) + poly(shld, 2) + htype + log(sigs1),
  data = Highway, method = "exhaustive", nvmax = 11, nbest = 1)
reg_summary <- summary(regfit_full)
```

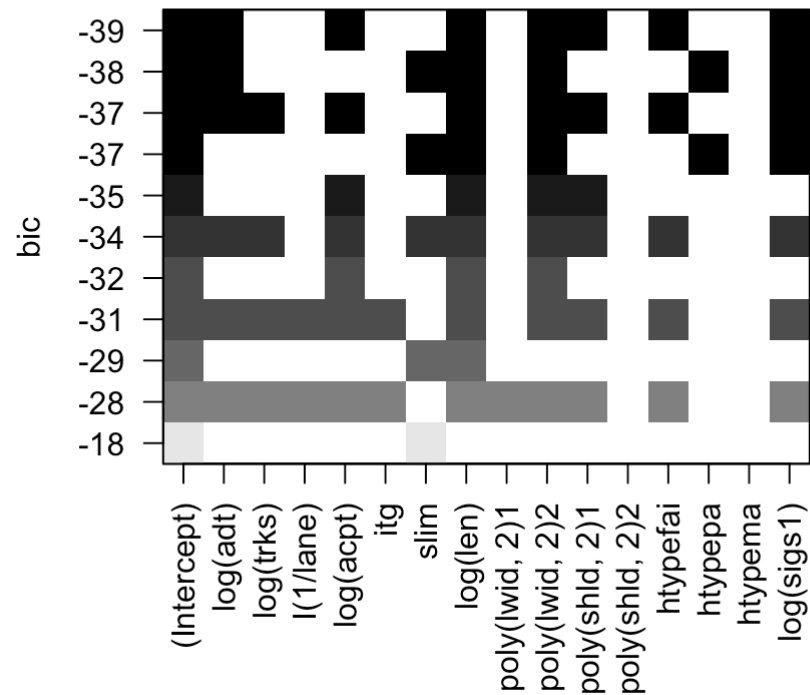
Investigating the results

```
plot(regfit_full, scale = "adjr2")
```



Investigating the results

```
plot(regfit_full, scale = "bic")
```



Another plot option

