Figures

```r
library(tidyverse)
library(MASS, exclude = "select")
library(marginaleffects)
library(broom)
library(survival)
```

Figure 1: Packages

Price	Beds	Baths	Size	Lot
82.0	3	3.0	1.454	2.50
92.5	3	1.0	1.329	0.42
92.5	3	2.0	1.790	0.73
144.0	3	1.5	1.968	0.34
118.0	3	1.0	1.500	0.50
92.0	3	2.0	1.815	0.00
100.5	3	2.0	1.608	0.58
114.0	4	2.0	2.720	2.50
197.5	4	2.5	2.268	1.00
72.0	3	2.0	1.442	0.22
185.0	4	3.5	2.220	0.12
115.0	4	2.0	1.401	1.10
87.0	3	1.5	1.740	0.25
179.0	4	2.0	3.100	0.28
162.0	4	1.5	2.044	0.85

Figure 2: Canton houses data (15 randomly chosen rows)

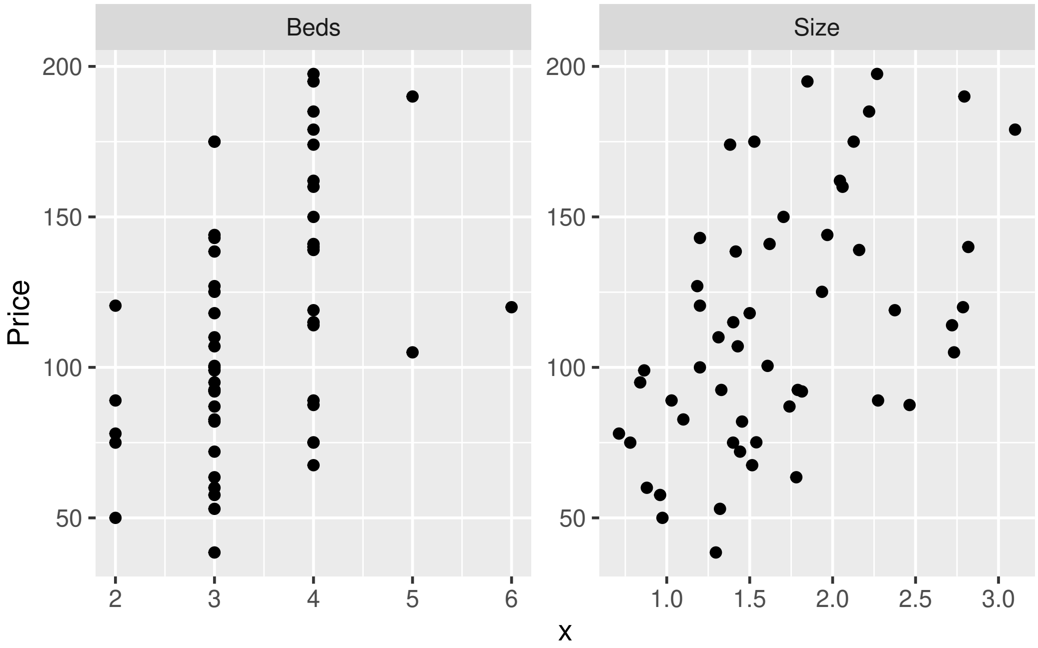

Figure 3: Canton houses scatterplots

```
boxcox(Price ~ Size + Beds, data = houses_ny)
```

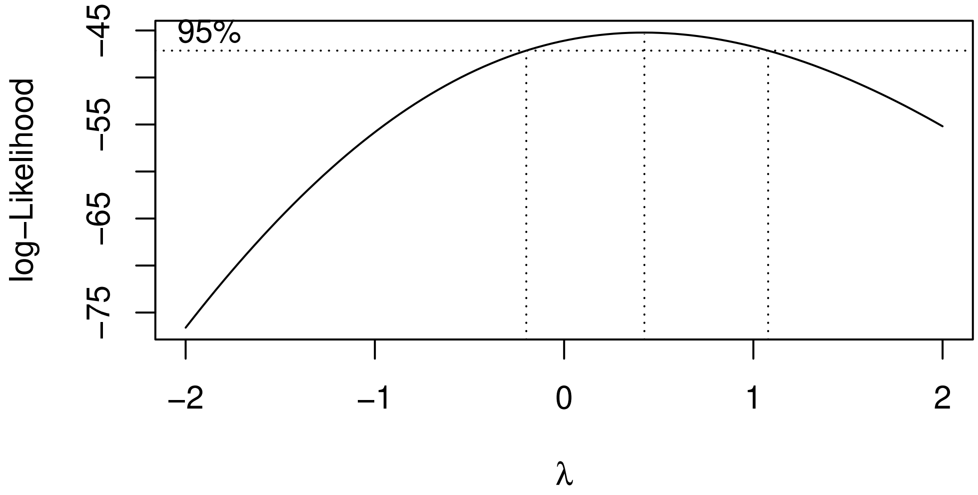

Figure 4: Code and output from houses data

```
houses.1 <- lm(Price ~ Size + Beds, data = houses_ny)
summary(houses.1)
```

```

Call:
lm(formula = Price ~ Size + Beds, data = houses_ny)

Residuals:
    Min      1Q  Median      3Q     Max 
-61.493 -31.920   1.696  27.866  73.436 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   46.498     22.277   2.087    0.042 *
Size          31.169     12.617   2.470    0.017 *
Beds           4.367      9.515   0.459    0.648  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.21 on 50 degrees of freedom
Multiple R-squared:  0.2653,	Adjusted R-squared:  0.236 
F-statistic:  9.03 on 2 and 50 DF,  p-value: 0.0004489
```

Figure 5: Regression 1 for Canton houses data

```
houses.2 <- lm(Price ~ Beds, data = houses_ny)
summary(houses.2)
```

```

Call:
lm(formula = Price ~ Beds, data = houses_ny)

Residuals:
    Min      1Q  Median      3Q     Max 
-66.453 -32.953  -5.048  33.142  70.642 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)   39.239     23.161   1.694  0.09632 . 
Beds          21.905      6.644   3.297  0.00179 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.98 on 51 degrees of freedom
Multiple R-squared:  0.1757,	Adjusted R-squared:  0.1595 
F-statistic: 10.87 on 1 and 51 DF,  p-value: 0.001785
```

Figure 6: Regression 2 for Canton houses data

```
new <- tribble(
  ~Size, ~Beds,
  2.75, 5,
  2.75, 2
)
cbind(predictions(houses.1, new)) %>%
  select(Beds, Size, estimate, conf.low, conf.high) %>%
  mutate(conf.length = conf.high - conf.low)
```

```
# A tibble: 2 x 6
   Beds  Size estimate conf.low conf.high conf.length
  <dbl> <dbl>    <dbl>    <dbl>     <dbl>       <dbl>
1     5  2.75     154.     131.      177.        45.1
2     2  2.75     141.      90.9     191.       100.
```

Figure 7: Canton houses data: predictions (values rounded to 3 significant digits)

```
houses_ny %>%
  summarize(mean_beds = mean(Beds),
            mean_size = mean(Size))
```

```
# A tibble: 1 x 2
  mean_beds mean_size
      <dbl>     <dbl>
1      3.40      1.68
```

```
ggplot(houses_ny, aes(x = Beds, y = Size)) + geom_point()
```

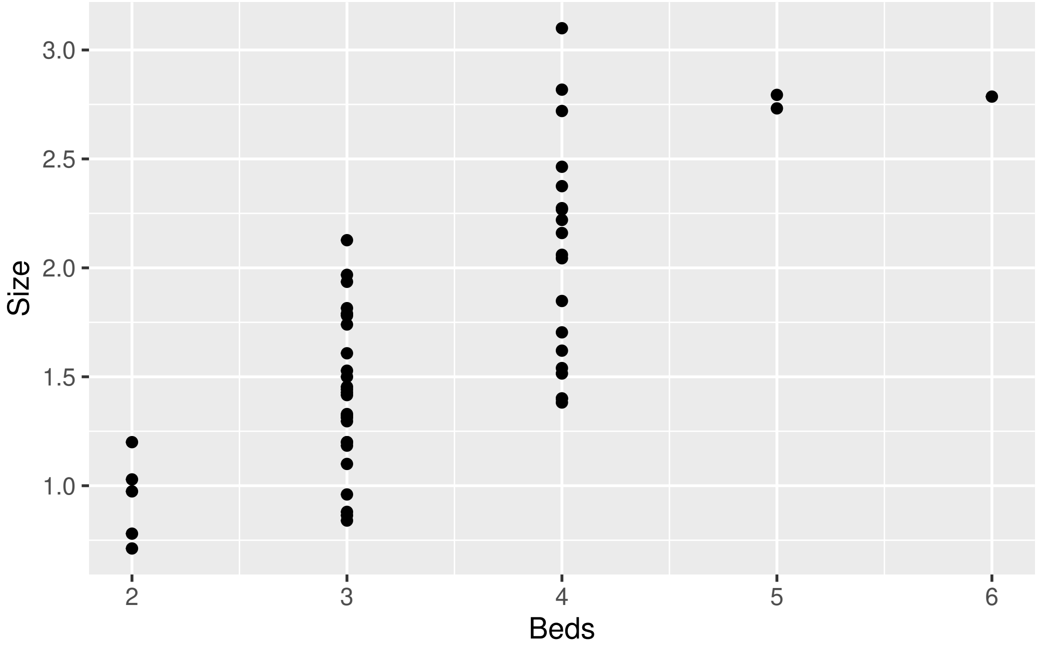

Figure 8: Canton houses data: further analysis

```
# A tibble: 15 x 3
    temp  male female
   <dbl> <int>  <int>
 1  27.2     1      9
 2  27.2     0      8
 3  27.2     1      8
 4  27.7     7      3
 5  27.7     4      2
 6  27.7     6      2
 7  28.3    13      0
 8  28.3     6      3
 9  28.3     7      1
10  28.4     7      3
11  28.4     5      3
12  28.4     7      2
13  29.9    10      1
14  29.9     8      0
15  29.9     9      0
```

Figure 9: Turtle hatch data, in dataframe `turtle` (all)

```
turtle.1 <- glm(cbind(female, male) ~ temp, data = turtle,
                family = "binomial")
summary(turtle.1)
```

```
Call:
glm(formula = cbind(female, male) ~ temp, family = "binomial",
    data = turtle)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  61.3183    12.0224   5.100 3.39e-07 ***
temp         -2.2110     0.4309  -5.132 2.87e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 74.508  on 14  degrees of freedom
Residual deviance: 24.942  on 13  degrees of freedom
AIC: 53.836

Number of Fisher Scoring iterations: 5
```

Figure 10: Turtle hatch data logistic regression

```
plot_predictions(model = turtle.1, condition = "temp")
```

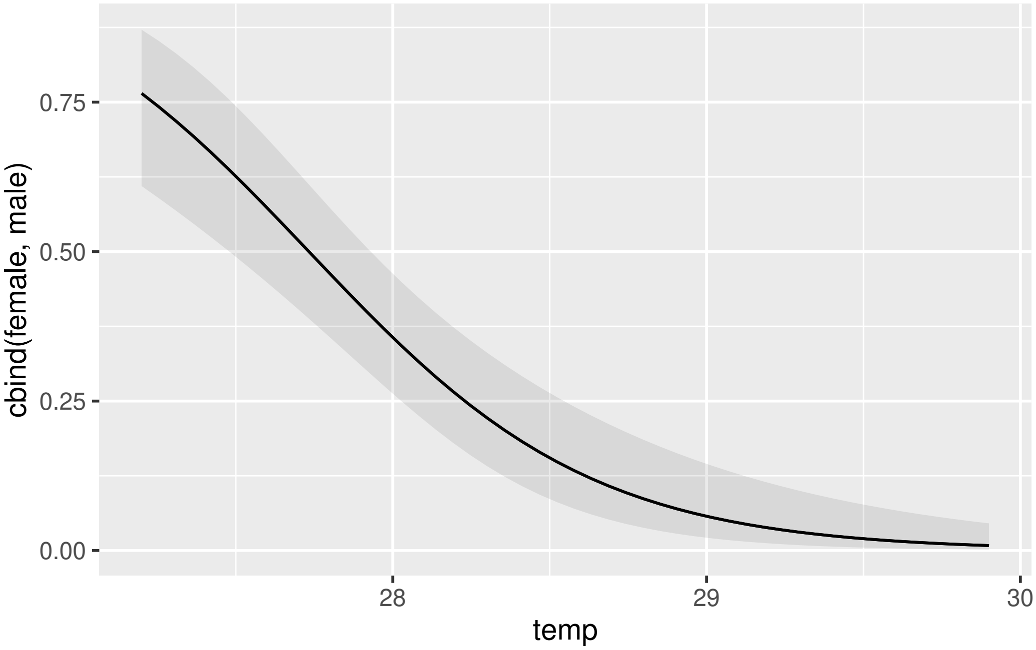

Figure 11: Turtle hatch data plot

```
# A tibble: 15 x 4
   sex     year use       n
   <chr>  <dbl> <fct> <int>
 1 female    80 <1m      30
 2 male      77 never    88
 3 female    80 >1m      15
 4 female    80 never    75
 5 female    77 never   106
 6 female    78 >1m       8
 7 female    79 >1m      14
 8 male      78 never    75
 9 male      76 never   104
10 male      79 <1m      20
11 male      78 <1m      20
12 male      79 never    70
13 female    78 never    91
14 female    76 >1m       1
15 male      80 >1m      32
```

Figure 12: Marijuana use data (15 randomly chosen rows)

```
potuse %>% count(use)
```

```
# A tibble: 3 x 2
  use       n
  <fct> <int>
1 never    10
2 <1m      10
3 >1m      10
```

Figure 13: Marijuana use summary

```
potuse.1 <- polr(use ~ sex + year, data = potuse, weights = n)
```

Figure 14: Marijuana use model

```
drop1(potuse.1, test = "Chisq")
```

	Df	AIC	LRT	Pr(>Chi)
	NA	1675.796	NA	NA
sex	1	1694.766	20.96921	4.7e-06
year	1	1771.661	97.86416	0.0e+00

Figure 15: Marijuana use model output

```
new <- datagrid(model = potuse.1,
                year = c(76, 78, 80),
                sex = c("male", "female"))
cbind(predictions(potuse.1, newdata = new)) %>%
  select(year, sex, group, estimate) %>%
  mutate(estimate = round(estimate, 3)) %>%
  pivot_wider(names_from = group, values_from = estimate)
```

```
Re-fitting to get Hessian

# A tibble: 6 x 5
   year sex    never `<1m` `>1m`
  <dbl> <chr>  <dbl> <dbl> <dbl>
1    76 male   0.858 0.09  0.053
2    76 female 0.918 0.053 0.029
3    78 male   0.697 0.176 0.127
4    78 female 0.811 0.117 0.073
5    80 male   0.467 0.256 0.277
6    80 female 0.62  0.21  0.17
```

Note: the predictions are shown to three decimal places. If only two decimal places are shown, the third one is zero.

Figure 16: Marijuana use: predictions

```
plot_predictions(model = potuse.1, condition = c("year", "group", "sex"))
```

```
Re-fitting to get Hessian
```

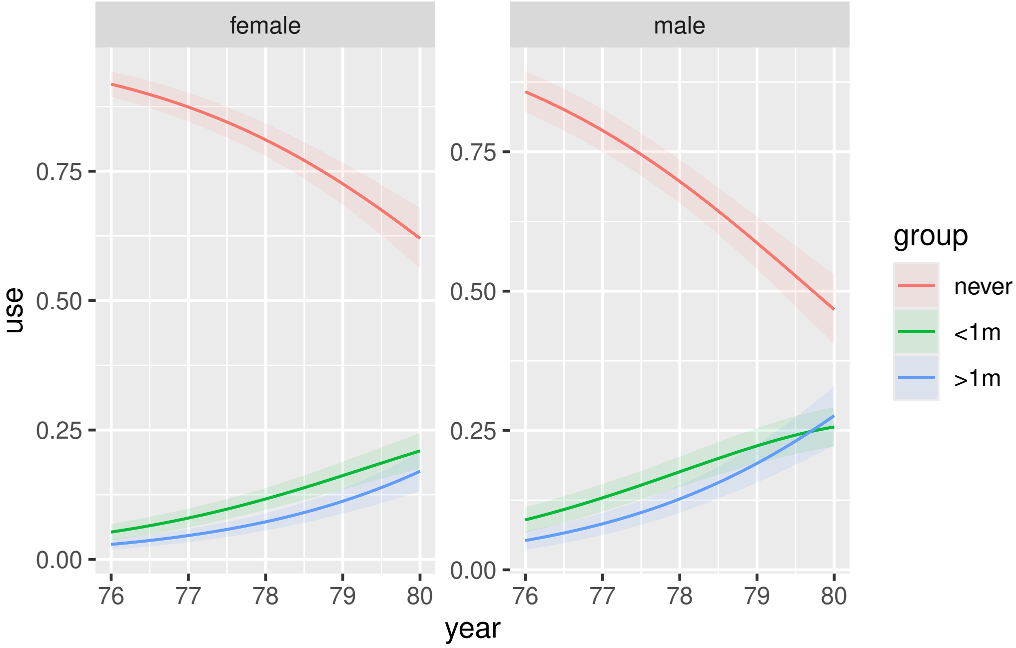

Figure 17: Marijuana use: plot

```
lungcancer
```

```
# A tibble: 14 x 3
    time  cens group
   <dbl> <dbl> <fct>
 1   257     0 newdrug
 2   476     0 newdrug
 3   355     1 newdrug
 4  1779     1 newdrug
 5   355     0 newdrug
 6   191     1 control
 7   563     1 control
 8   242     1 control
 9   285     1 control
10    16     1 control
11    16     1 control
12    16     1 control
13   257     1 control
14    16     1 control
```

Figure 18: Lung cancer data

```
with(lungcancer, Surv(time, cens == 1))
```

```
 [1]  257+  476+  355  1779   355+  191   563   242   285    16    16    16
[13]  257    16
```

Figure 19: Lung cancer code and its output

```
lungcancer.1 <- coxph(Surv(time, cens == 1) ~ group, data = lungcancer)
summary(lungcancer.1)
```

```
Call:
coxph(formula = Surv(time, cens == 1) ~ group, data = lungcancer)

  n= 14, number of events= 11 

                 coef exp(coef) se(coef)      z Pr(>|z|)  
groupnewdrug -2.45904   0.08552  1.07581 -2.286   0.0223 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

             exp(coef) exp(-coef) lower .95 upper .95
groupnewdrug   0.08552      11.69   0.01038    0.7043

Concordance= 0.764  (se = 0.064 )
Likelihood ratio test= 8.84  on 1 df,   p=0.003
Wald test            = 5.22  on 1 df,   p=0.02
Score (logrank) test = 7.82  on 1 df,   p=0.005
```

Figure 20: Lung cancer Cox model

```
plot_predictions(lungcancer.1, condition = c("time", "group"),
                 type = "survival")
```

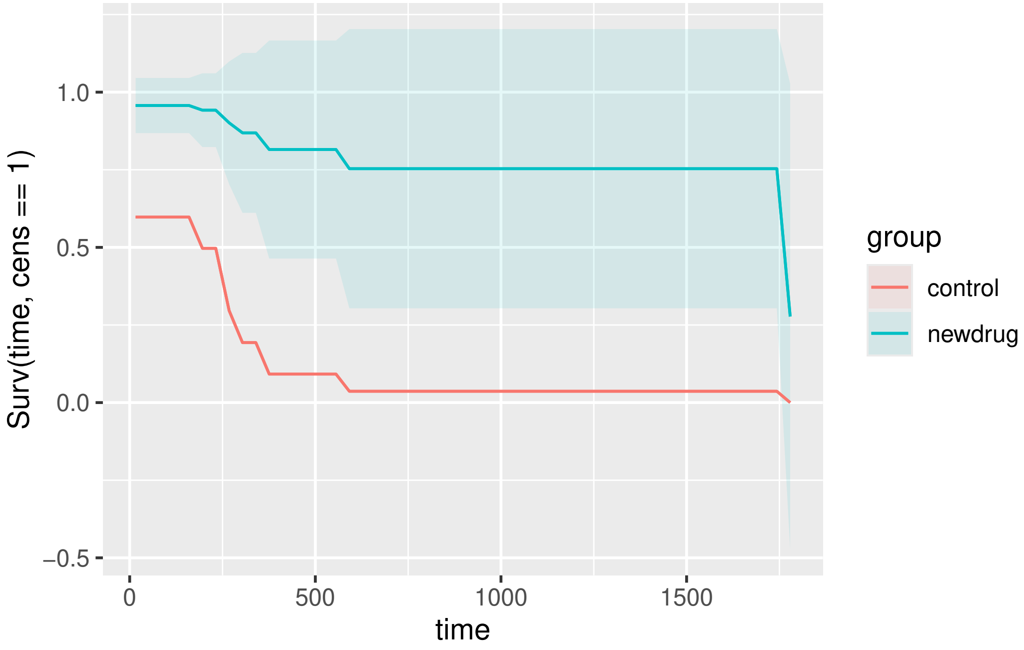

Figure 21: Predictions from lung cancer Cox model

```
lungcancer.1 %>% augment(lungcancer) %>%
  ggplot(aes(x = .fitted, y = .resid)) + geom_point()
```

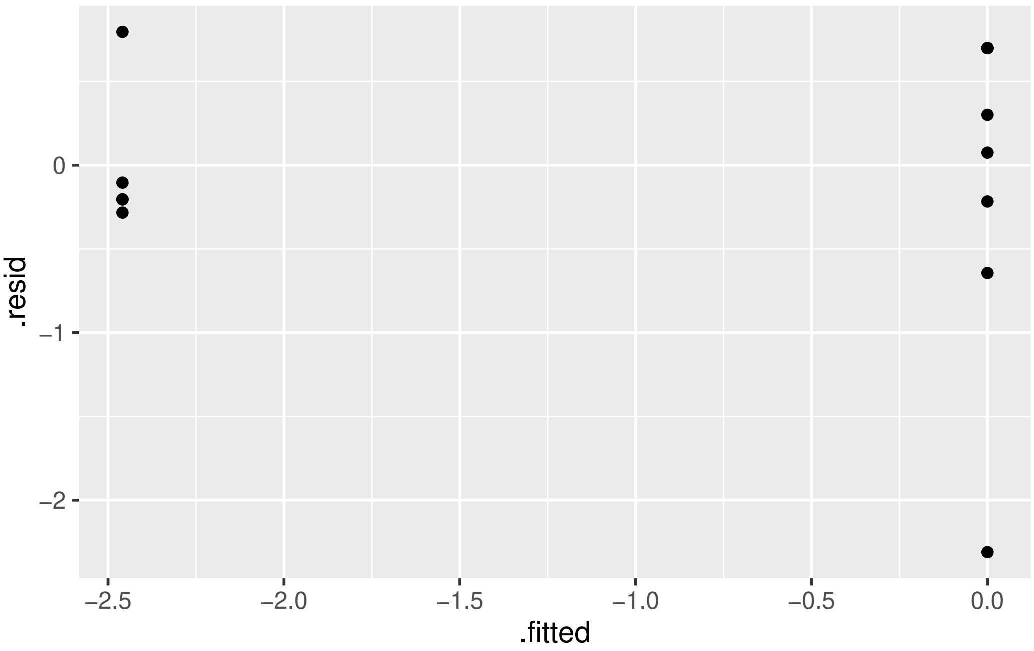

Figure 22: Another plot from lung cancer Cox model

group	attempts
Group2	10
Group1	3
Group3	15
Group2	13
Group1	4
Group2	11
Group1	7
Group2	9
Group3	18
Group1	7

Figure 23: Shock data (10 randomly chosen rows)

```
Call:
lm(formula = attempts ~ group, data = Shock)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.4444 -1.4444  0.1111  1.1111  3.5556 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    10.4074     0.3889  26.762  < 2e-16 ***
groupc_any     -4.5185     0.5500  -8.216 1.96e-08 ***
groupc_med_sev -2.2222     0.4763  -4.666 9.72e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.021 on 24 degrees of freedom
Multiple R-squared:  0.7881,	Adjusted R-squared:  0.7705 
F-statistic: 44.63 on 2 and 24 DF,  p-value: 8.188e-09
```

Figure 24: Shock data analysis

treatment	rep	weight1	feed	weight2	gain
T2	R1	26	699	194	168
T1	R9	28	632	192	164
T1	R10	26	637	184	158
T5	R6	36	678	196	160
T3	R5	32	607	185	153
T5	R3	32	722	205	173
T4	R3	32	733	218	186
T3	R7	30	637	190	160
T5	R7	30	763	230	200
T5	R9	32	710	216	184

Figure 25: Pigs data (10 randomly chosen rows)

```
ggplot(crampton.pig, aes(x = feed, y = gain, colour = treatment)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

```
`geom_smooth()` using formula = 'y ~ x'
```

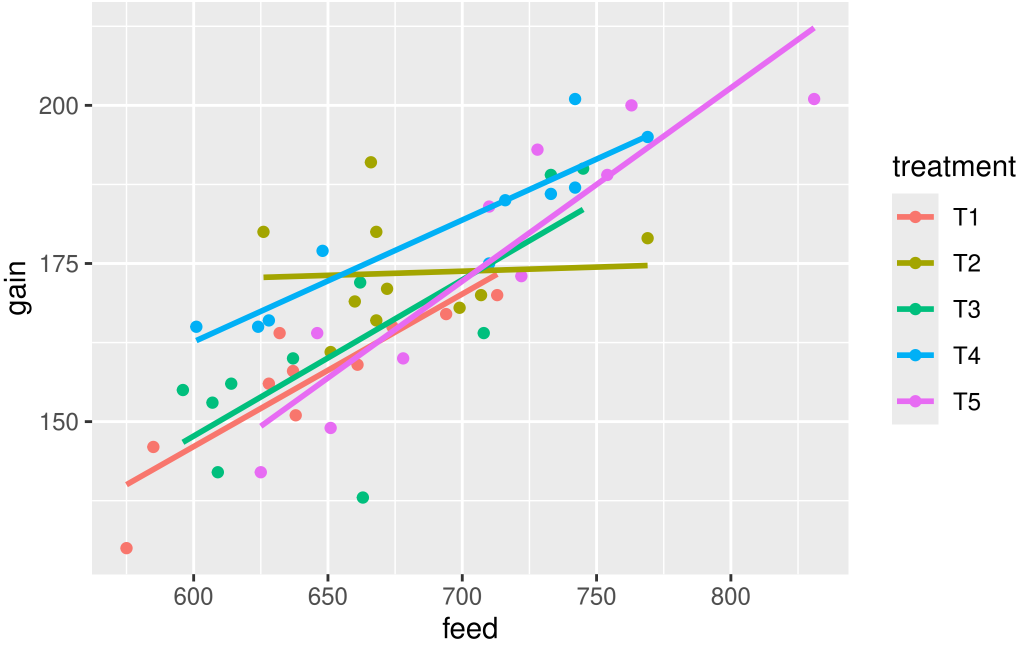

Figure 26: Graph of pigs data

```
pig.1 <- lm(gain ~ feed * treatment, data = crampton.pig)
drop1(pig.1, test = "F")
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
	NA	NA	3024.615	225.1258	NA	NA
feed:treatment	4	899.1275	3923.742	230.1389	2.972701	0.030638

Figure 27: ANCOVA of pigs data

```
summary(pig.1)
```

```
Call:
lm(formula = gain ~ feed * treatment, data = crampton.pig)

Residuals:
     Min       1Q   Median       3Q      Max 
-25.2835  -5.1686  -0.0565   6.1270  17.6647 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        1.342383  42.798105   0.031 0.975134    
feed               0.241196   0.066350   3.635 0.000784 ***
treatmentT2      163.288907  66.193398   2.467 0.018011 *  
treatmentT3       -1.860747  55.241490  -0.034 0.973297    
treatmentT4       45.776065  54.342512   0.842 0.404594    
treatmentT5      -43.145509  53.945914  -0.800 0.428556    
feed:treatmentT2  -0.228126   0.099615  -2.290 0.027368 *  
feed:treatmentT3   0.005866   0.084898   0.069 0.945258    
feed:treatmentT4  -0.048686   0.082056  -0.593 0.556298    
feed:treatmentT5   0.064521   0.080759   0.799 0.429050    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.696 on 40 degrees of freedom
Multiple R-squared:  0.7857,	Adjusted R-squared:  0.7375 
F-statistic:  16.3 on 9 and 40 DF,  p-value: 8.356e-11
```

Figure 28: Summary output from ANCOVA