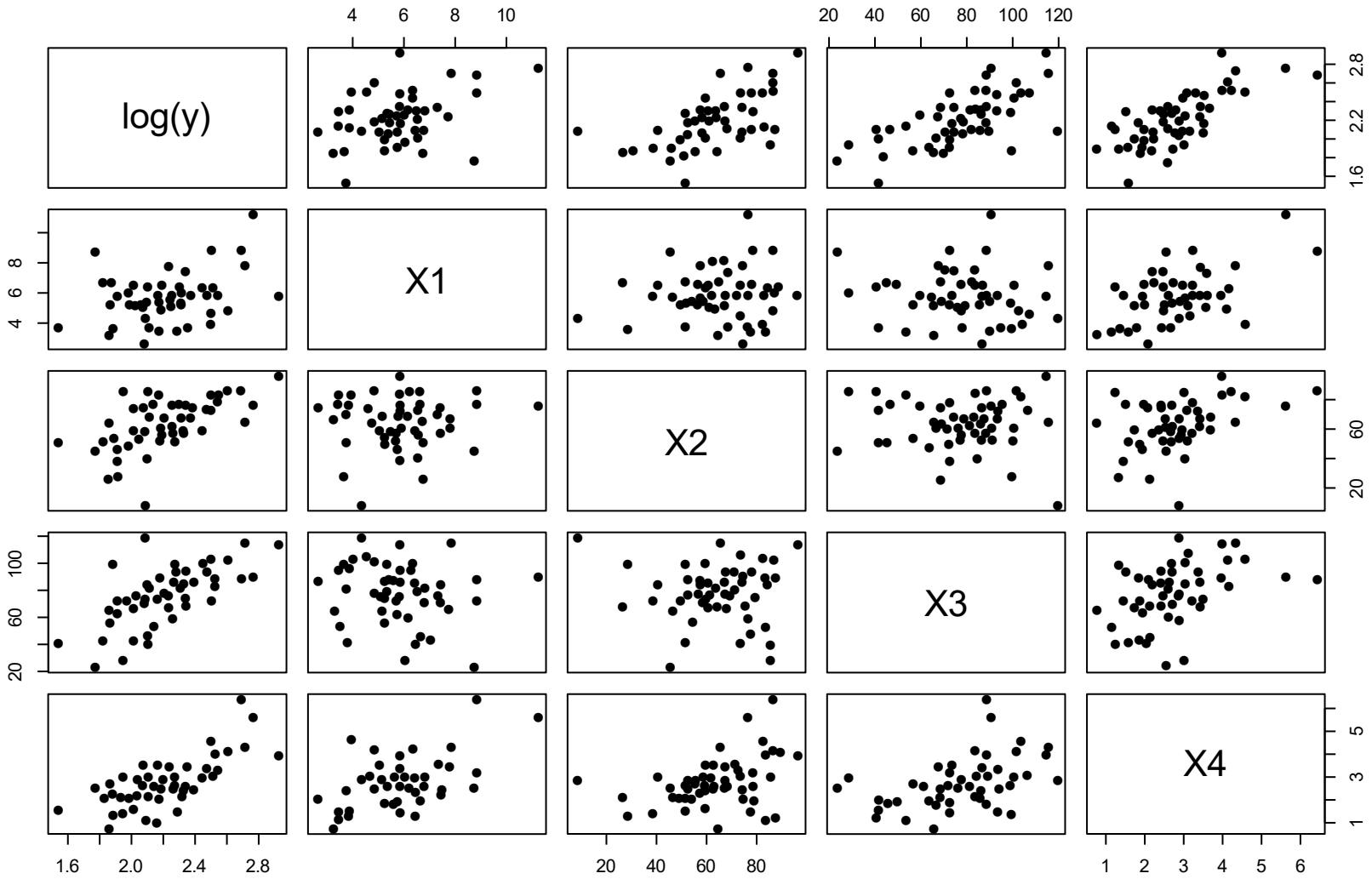
Example

A hospital surgical unit was interested in predicting survival in patients undergoing a particular type of liver operation. A random selection of 54 patients was available for analysis. The potential predictors are

X_1 blood clotting score
X_2 prognostic index, includes age of patient
X_3 enzyme function test score
X_4 liver function test score

The number of candidate regressors is small. We can easily eplore the different models.


log(y)
X1
X2
X3
X4
4 6 8 10
20 40 60 80 100 120
1.6 2.2 2.8
4 6 8
20 60
20 60 100
1 3 5
1.6 2.0 2.4 2.8
20 40 60 80
1 2 3 4 5 6

Model	AIC	mse	R^2	R^2_{adj}
X_1, X_2, X_3, X_4	−324.71	0.00224	0.972*	0.970
X_1, X_2, X_3	−326.67*	0.00220*	0.972	0.971*
X_1, X_2, X_4	−189.60	0.0278	0.65	0.63
X_1, X_2	−166.04	0.0438	0.44	0.42
X_1, X_3, X_4	−201.50	0.0223	0.72	0.70
X_1, X_3	−190.96	0.0276	0.65	0.63
X_1, X_4	−175.44	0.0368	0.53	0.51
X_1	−143.82	0.0672	0.12	0.10
X_2, X_3, X_4	−248.73	0.0093	0.88	0.88
X_2, X_3	−225.45	0.046	0.81	0.81
X_2, X_4	−191.54	0.0273	0.65	0.64
X_2	−160.30	0.0495	0.35	0.34
X_3, X_4	−197.56	0.0244	0.69	0.67
X_3	−168.45	0.0426	0.44	0.43
X_4	−177.38	0.0361	0.53	0.52

Selection using AIC, MSE and adjusted R^2 all yield the same

Stepwise regression methods

- Evaluating all possible regressions can be burdensome computationally and for the analyst.
- An alternative might be to compare a scientifically meaningful subset of models.
- When that's not possible or for some reason, not desired, procedures have been developed to evaluate only a small subset of regression models by either adding or deleting regressors one at a time. These fit into 3 categories
  - forward selection
  - backward elimination
  - stepwise regression

1. Begin with the assumption that there are no regressors in the model.

2. Check models with all possible regressors added individually.

3. Add the regressor that most changes your criterion in the correct direction. Go back to 2.

4. If none of the regressors have a positive effect on your criterion, stop with the regressors you have. This is your final model.

Forward selection example

```
Forward selection
Single term additions

Model:
surg$logY ~ -1 + 1
        Df Sum of Sq      RSS      AIC
<none>                  3.973 -138.914
surg$X1  1     0.477    3.496 -143.817
surg$X2  1     1.396    2.576 -160.303
surg$X3  1     1.758    2.215 -168.455
surg$X4  1     2.095    1.878 -177.385
Single term additions

Model:
surg$logY ~ surg$X4
        Df Sum of Sq      RSS      AIC
<none>                  1.878 -177.385
surg$X1  1     0.002    1.876 -175.437
surg$X2  1     0.485    1.392 -191.539
```

```
surg$X3  1     0.632    1.245 -197.558
Single term additions

Model:
surg$logY ~ surg$X4 + surg$X3
        Df Sum of Sq      RSS      AIC
<none>                   1.245 -197.558
surg$X1  1     0.130    1.116 -201.498
surg$X2  1     0.780    0.465 -248.730
Single term additions

Model:
surg$logY ~ surg$X4 + surg$X3 + surg$X2
        Df Sum of Sq     RSS     AIC
<none>                  0.47 -248.73
surg$X1  1      0.36    0.11 -324.71
```

Note: RSS=residual sums of squares (SSE)

Model selected by forward selection using AIC:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.4888	0.0502	9.73	0.0000
X4	0.0019	0.0097	0.20	0.8436
X3	0.0095	0.0004	23.91	0.0000
X2	0.0093	0.0004	21.19	0.0000
X1	0.0685	0.0054	12.60	0.0000

Backward elimination example

```
Backward elimination
Single term deletions

Model:
surg$logY ~ surg$X4 + surg$X3 + surg$X2 + surg$X1
         Df Sum of Sq     RSS     AIC
<none>                    0.11 -324.71
surg$X1   1      0.36    0.47 -248.73
surg$X2   1      1.01    1.12 -201.50
surg$X3   1      1.28    1.39 -189.60
surg$X4   1  8.81e-05    0.11 -326.67

Model:
surg$logY ~ surg$X3 + surg$X2 + surg$X1
         Df Sum of Sq     RSS     AIC
<none>                    0.11 -326.67
surg$X1   1      0.63    0.74 -225.45
surg$X2   1      1.30    1.41 -190.96
surg$X3   1      2.12    2.23 -166.04
```

Model selected by backward elimination

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.4836	0.0426	11.34	0.0000
X3	0.0095	0.0003	31.08	0.0000
X2	0.0093	0.0004	24.30	0.0000
X1	0.0692	0.0041	16.98	0.0000

Example using Stepwise regression

```
Start with no regressors
Start:  AIC= -138.91
 surg$logY ~ -1 + 1

           Df Sum of Sq      RSS      AIC
+ surg$X4   1     2.095    1.878 -177.385
+ surg$X3   1     1.758    2.215 -168.455
+ surg$X2   1     1.396    2.576 -160.303
+ surg$X1   1     0.477    3.496 -143.817
<none>                     3.973 -138.914

Step:  AIC= -177.38
 surg$logY ~ surg$X4

           Df Sum of Sq      RSS      AIC
+ surg$X3   1     0.632    1.245 -197.558
+ surg$X2   1     0.485    1.392 -191.539
```

```
<none>                      1.878 -177.385
+ surg$X1  1     0.002    1.876 -175.437
- surg$X4  1     2.095    3.973 -138.914

Step:  AIC= -197.56
 surg$logY ~ surg$X4 + surg$X3

          Df Sum of Sq      RSS      AIC
+ surg$X2  1     0.780    0.465 -248.730
+ surg$X1  1     0.130    1.116 -201.498
<none>                    1.245 -197.558
- surg$X3  1     0.632    1.878 -177.385
- surg$X4  1     0.970    2.215 -168.455

Step:  AIC= -248.73
 surg$logY ~ surg$X4 + surg$X3 + surg$X2

          Df Sum of Sq     RSS     AIC
+ surg$X1  1      0.36    0.11 -324.71
<none>                    0.47 -248.73
- surg$X4  1      0.28    0.74 -225.45
```

```
- surg$X2  1      0.78    1.25 -197.56
- surg$X3  1      0.93    1.39 -191.54

Step:  AIC= -324.71
 surg$logY ~ surg$X4 + surg$X3 + surg$X2 + surg$X1

          Df Sum of Sq     RSS     AIC
- surg$X4  1  8.81e-05    0.11 -326.67
<none>                    0.11 -324.71
- surg$X1  1      0.36    0.47 -248.73
- surg$X2  1      1.01    1.12 -201.50
- surg$X3  1      1.28    1.39 -189.60

Step:  AIC= -326.67
 surg$logY ~ surg$X3 + surg$X2 + surg$X1

          Df Sum of Sq     RSS     AIC
<none>                    0.11 -326.67
+ surg$X4  1  8.81e-05    0.11 -324.71
- surg$X1  1      0.63    0.74 -225.45
- surg$X2  1      1.30    1.41 -190.96
```

```
- surg$X3  1      2.12    2.23 -166.04

Call:
lm(formula = surg$logY ~ surg$X3 + surg$X2 + surg$X1)

Coefficients:
(Intercept)      surg$X3      surg$X2      surg$X1
   0.483621     0.009524     0.009295     0.069225

Start with full model
Start:  AIC= -324.71
 surg$logY ~ surg$X4 + surg$X3 + surg$X2 + surg$X1

          Df Sum of Sq     RSS     AIC
- surg$X4  1  8.81e-05    0.11 -326.67
<none>                    0.11 -324.71
- surg$X1  1      0.36    0.47 -248.73
- surg$X2  1      1.01    1.12 -201.50
- surg$X3  1      1.28    1.39 -189.60

Step:  AIC= -326.67
```

```
 surg$logY ~ surg$X3 + surg$X2 + surg$X1

          Df Sum of Sq     RSS     AIC
<none>                    0.11 -326.67
- surg$X1  1      0.63    0.74 -225.45
- surg$X2  1      1.30    1.41 -190.96
- surg$X3  1      2.12    2.23 -166.04

Call:
lm(formula = surg$logY ~ surg$X3 + surg$X2 + surg$X1)

Coefficients:
(Intercept)      surg$X3      surg$X2      surg$X1
   0.483621     0.009524     0.009295     0.069225
```