# Example on One-Way ANOVA

- 24 animals were randomly assigned to four different diets and
- The blood coagulation time was measured. Box et al. (1978).

```r
library(faraway)
data(coagulation, package="faraway")
head(coagulation)
```

```
##    coag diet
## 1    62    A
## 2    60    A
## 3    63    A
## 4    59    A
## 5    63    B
## 6    67    B
```

```r
par(mfrow=c(1,2))
plot(coag ~ diet, coagulation,ylab="coagulation time")
stripchart(coag ~ diet, coagulation, vertical=TRUE, method="stack",
           xlab="diet",ylab="coagulation time")
```
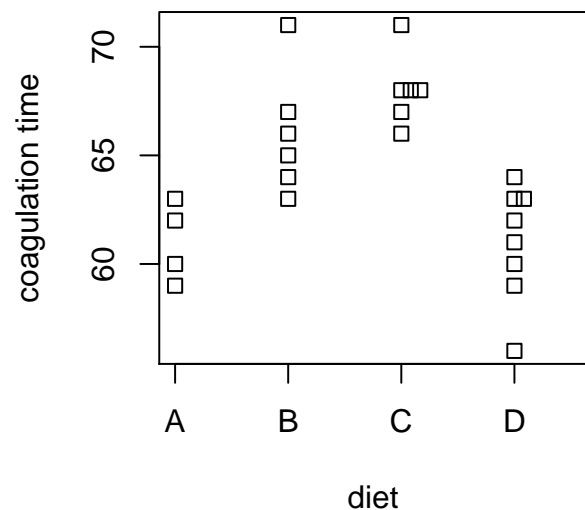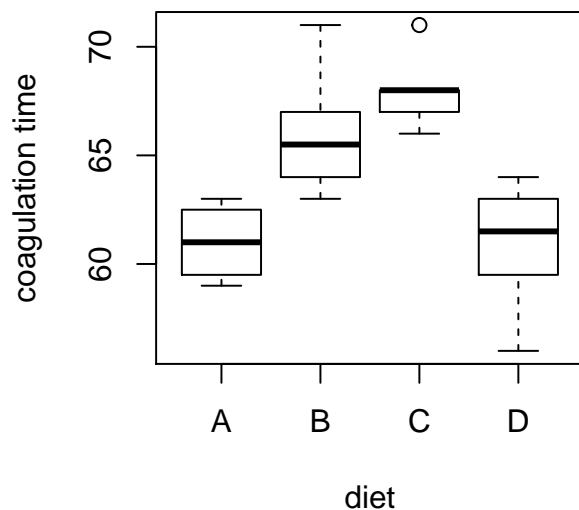
```
par(mfrow=c(1,1))
```

- Left: boxplot.
- Right: stripchart. (1-dim scatterplot, an alternative to boxplots when sample sizes are small.)
- Median and upper quartile of diet C are the same.
- There are ties in diets C and D.

ANOVA code version 1

```
lmodi <- lm(coag ~ diet -1, coagulation)
summary(lmodi)$coefficients
```

```
##        Estimate Std. Error  t value     Pr(>|t|)
## dietA       61  1.1832160 51.55441 9.547815e-23
## dietB       66  0.9660918 68.31649 3.532325e-25
## dietC       68  0.9660918 70.38669 1.948886e-25
## dietD       61  0.8366600 72.90895 9.663048e-26
```

```
lmnull <- lm(coag ~ 1, coagulation)
anova(lmnull,lmodi)
```

```
## Analysis of Variance Table
##
## Model 1: coag ~ 1
## Model 2: coag ~ diet - 1
##   Res.Df RSS Df Sum of Sq      F    Pr(>F)
## 1     23 340
## 2     20 112  3       228 13.571 4.658e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We see that there is indeed a difference in the levels.

ANOVA code version 2

```
lmod <- lm(coag ~ diet, coagulation)
summary(lmod)$coefficients
```

```
##                  Estimate Std. Error      t value      Pr(>|t|)
## (Intercept) 6.100000e+01   1.183216 5.155441e+01 9.547815e-23
## dietB       5.000000e+00   1.527525 3.273268e+00 3.802505e-03
## dietC       7.000000e+00   1.527525 4.582576e+00 1.805132e-04
## dietD       2.991428e-15   1.449138 2.064281e-15 1.000000e+00
```

```
anova(lmod)
```

```
## Analysis of Variance Table
##
## Response: coag
##           Df Sum Sq Mean Sq F value    Pr(>F)
## diet       3    228    76.0  13.571 4.658e-05 ***
## Residuals 20    112     5.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note

```
anova(lmnull, lmod) #This is also ok.
```

```
## Analysis of Variance Table
##
## Model 1: coag ~ 1
## Model 2: coag ~ diet
##   Res.Df RSS Df Sum of Sq      F    Pr(>F)
## 1     23 340
## 2     20 112  3       228 13.571 4.658e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lmodi) #This is incorrect
```

Analysis of Variance Table
##
Response: coag
Df Sum Sq Mean Sq F value   Pr(>F)
diet   4 98532 24633.0 4398.8 < 2.2e-16 ***
Residuals 20   112   5.6

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1