Chapter 4 Section 3.2

Dr Richard Wilkinson

8 February 2016

```
library(lme4)
```

```
## Loading required package: Matrix
```

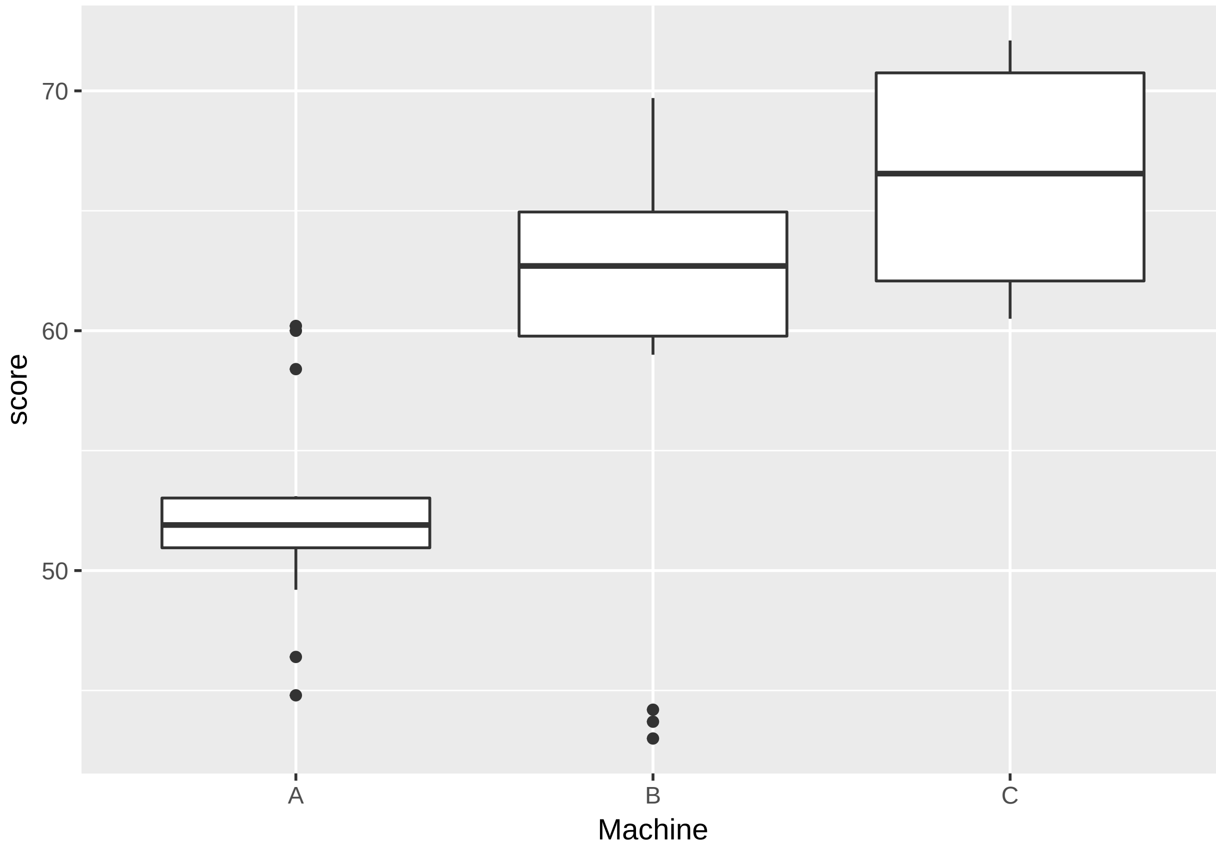
```
load('MAS473.RData')
```

Chapter 4 Section 3.2

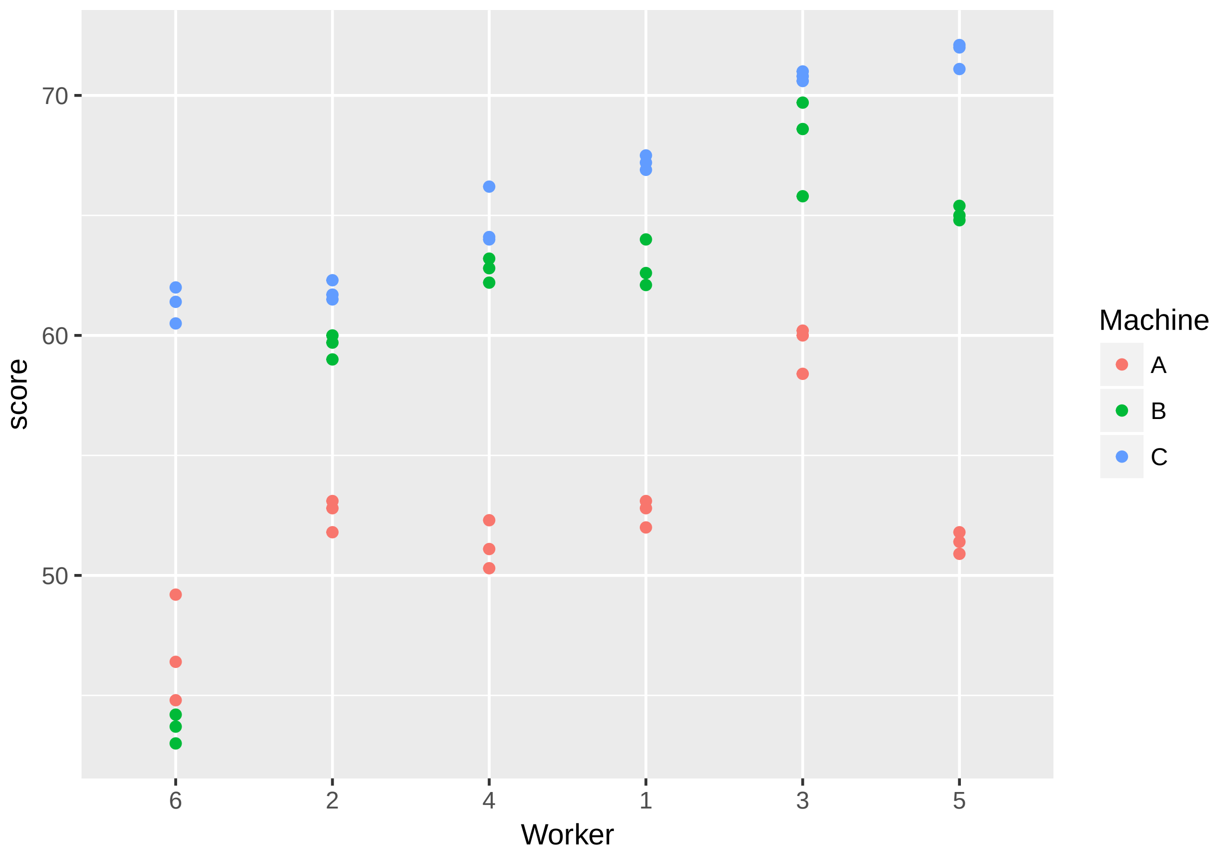
```
attach(Machines)



head(Machines)
```

```
##   Worker Machine score
## 1      1       A  52.0
## 2      1       A  52.8
## 3      1       A  53.1
## 4      2       A  51.8
## 5      2       A  52.8
## 6      2       A  53.1
```

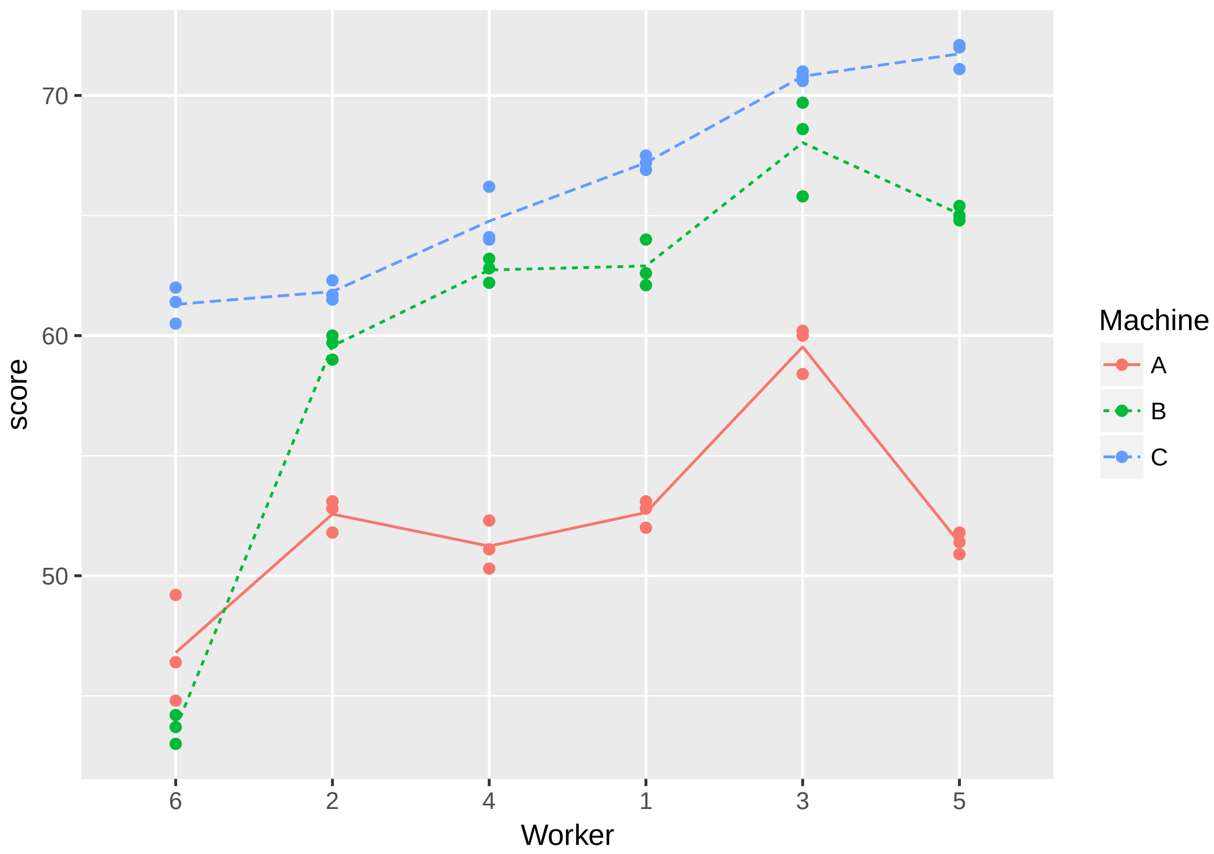
```
library(ggplot2)
qplot(Machine, score, geom='boxplot')
```



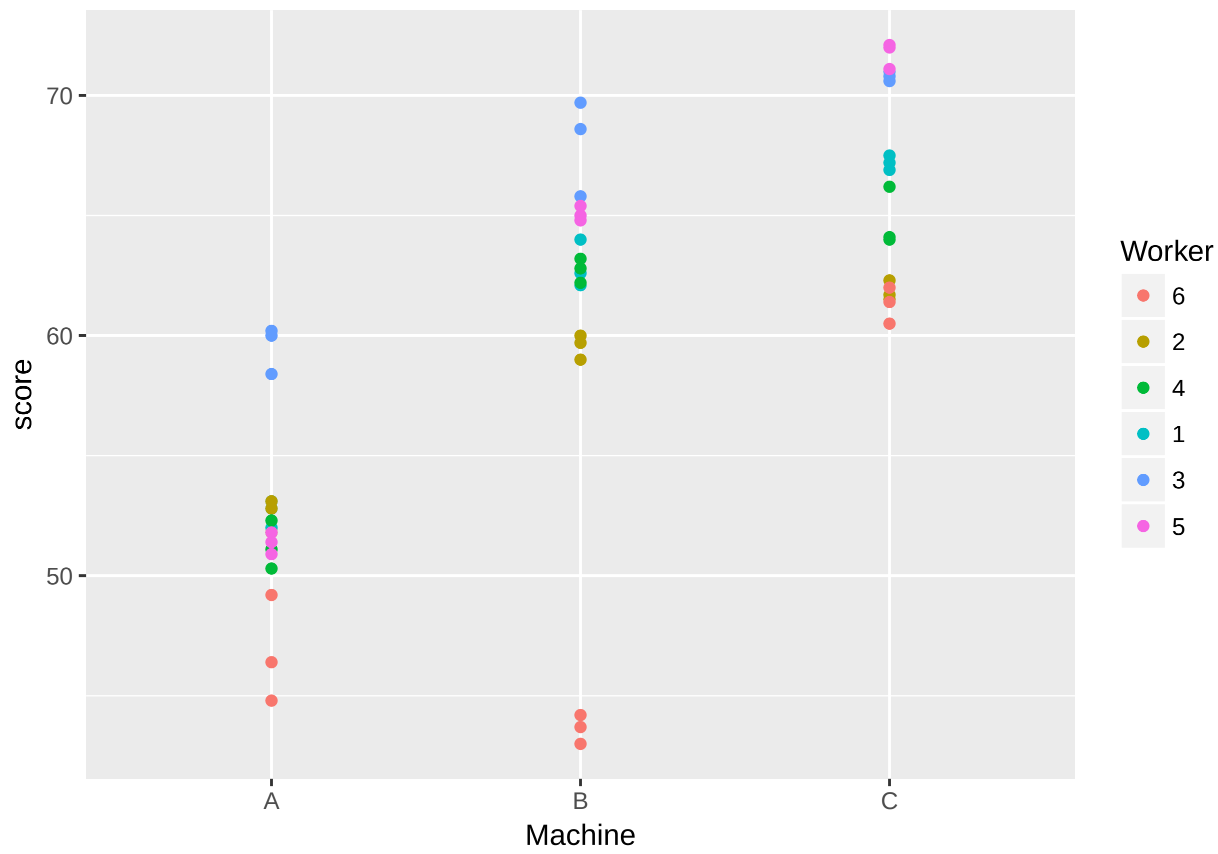

```
(q<-qplot(Worker, score, color=Machine))
```



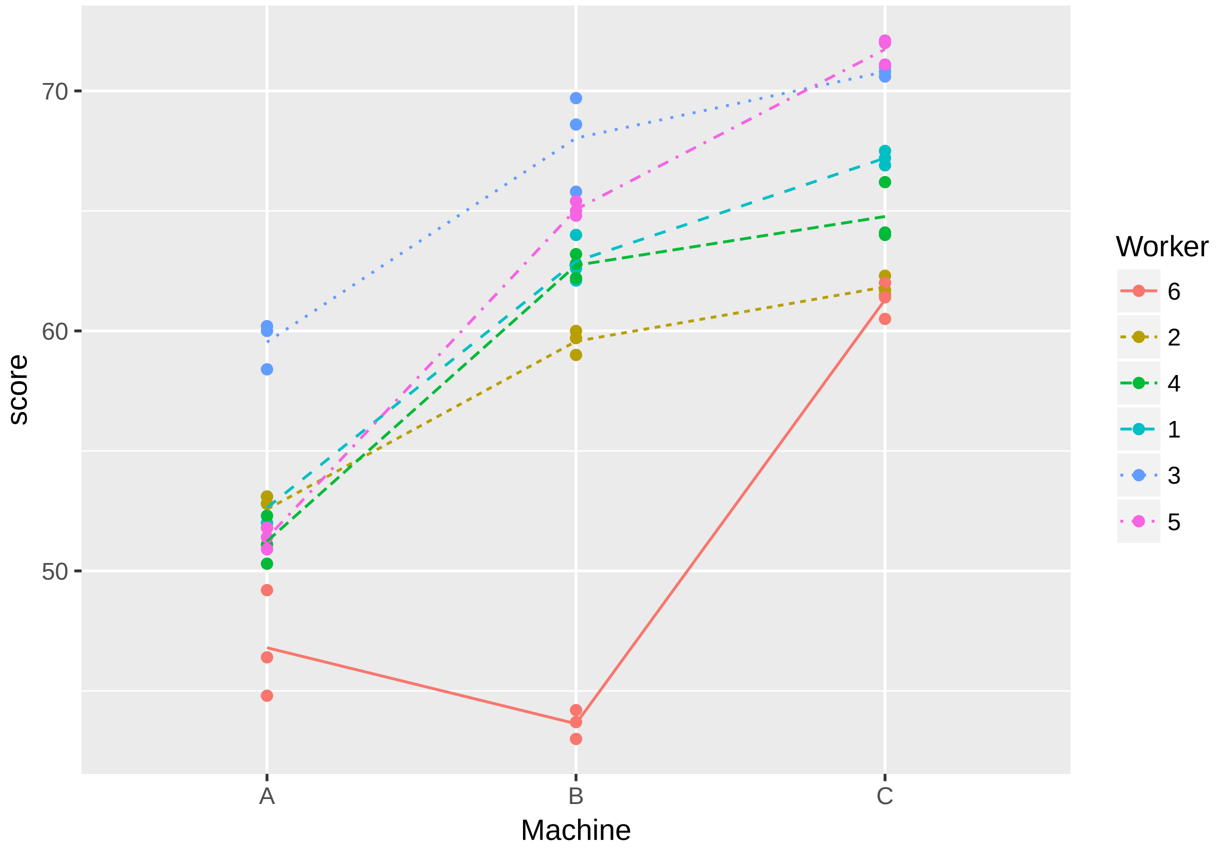

```r
q+stat_summary(fun.y = mean, geom = "line", aes(group = Machine, lty=Machine))
```




```
(g <- qplot(Machine, score, color=Worker))
```

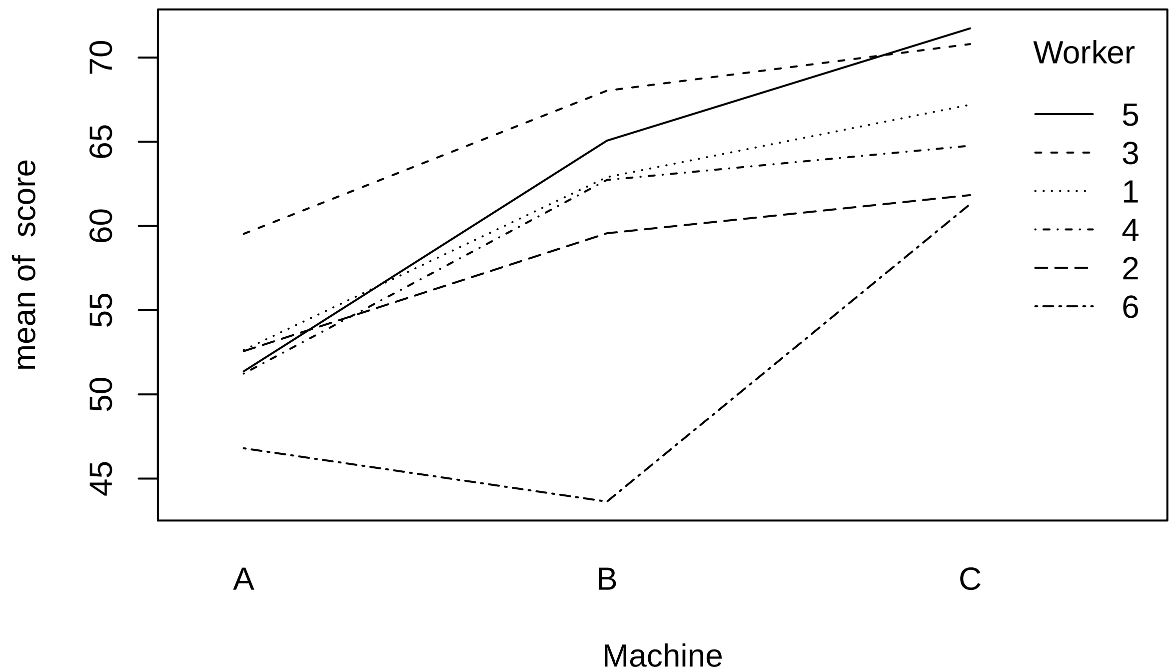



```
g + stat_summary(fun.y = mean, geom = "line", aes(group = Worker, lty=Worker))
```



But for once, these types of plots are easier using the built in interaction.plot command

```
interaction.plot(Machine,Worker,score)
```



Plot suggests significant machine effects, with C better than B better than A Suggests significant worker random effects, with worker 3 the best Also suggests interaction, looking at worker 6

```
(fm1<-lmer(score~Machine-1+(1|Worker/Machine),data=Machines))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ Machine - 1 + (1 | Worker/Machine)
##    Data: Machines
## REML criterion at convergence: 215.6876
## Random effects:
##  Groups         Name        Std.Dev.
##  Machine:Worker (Intercept) 3.7295
##  Worker         (Intercept) 4.7811
##  Residual                   0.9616
## Number of obs: 54, groups:  Machine:Worker, 18; Worker, 6
## Fixed Effects:
## MachineA  MachineB  MachineC
##    52.36     60.32     66.27
```

```
summary(fm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ Machine - 1 + (1 | Worker/Machine)
##    Data: Machines
##
## REML criterion at convergence: 215.7
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.26959 -0.54847 -0.01071  0.43937  2.54006
##
## Random effects:
##  Groups         Name        Variance Std.Dev.
##  Machine:Worker (Intercept) 13.9095  3.7295
##  Worker         (Intercept) 22.8584  4.7811
##  Residual                    0.9246  0.9616
## Number of obs: 54, groups:  Machine:Worker, 18; Worker, 6
##
## Fixed effects:
##          Estimate Std. Error t value
## MachineA   52.356      2.486   21.06
## MachineB   60.322      2.486   24.27
## MachineC   66.272      2.486   26.66
##
## Correlation of Fixed Effects:
##          MachnA MachnB
## MachineB 0.617
## MachineC 0.617  0.617
```

Now consider the following

```
(fm3<-lmer(score~Machine-1+(Machine-1|Worker),data=Machines))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ Machine - 1 + (Machine - 1 | Worker)
##    Data: Machines
## REML criterion at convergence: 208.3112
## Random effects:
##  Groups   Name     Std.Dev. Corr
##  Worker   MachineA 4.0793
##           MachineB 8.6253   0.80
##           MachineC 4.3895   0.62 0.77
##  Residual          0.9616
## Number of obs: 54, groups:  Worker, 6
## Fixed Effects:
## MachineA  MachineB  MachineC
##    52.36     60.32     66.27
```

```
summary(fm3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ Machine - 1 + (Machine - 1 | Worker)
##    Data: Machines
##
## REML criterion at convergence: 208.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.39354 -0.51378  0.02691  0.47245  2.53339
##
## Random effects:
##  Groups   Name     Variance Std.Dev. Corr
##  Worker   MachineA 16.6405  4.0793
##           MachineB 74.3956  8.6253   0.80
```

```
##           MachineC 19.2675  4.3895   0.62 0.77
##  Residual           0.9246  0.9616
## Number of obs: 54, groups:  Worker, 6
##
## Fixed effects:
##          Estimate Std. Error t value
## MachineA   52.356      1.681   31.15
## MachineB   60.322      3.529   17.10
## MachineC   66.272      1.806   36.69
##
## Correlation of Fixed Effects:
##          MachnA MachnB
## MachineB 0.794
## MachineC 0.612  0.763
```

and compare

```
ranef(fm1)
```

```
## $`Machine:Worker`
##     (Intercept)
## A:6   1.9162737
## A:2   1.5525649
## A:4  -1.0393717
## A:1  -0.7501465
## A:3   1.7776121
## A:5  -3.4569326
## B:6  -8.9757118
## B:2   0.6068536
## B:4   2.4173662
## B:1   1.4999942
## B:3   2.2993839
## B:5   2.1521138
## C:6   2.4869615
## C:2  -2.9966326
## C:4  -1.4143951
## C:1  -0.1142371
## C:3  -0.8149413
## C:5   2.8532446
##
## $Worker
##   (Intercept)
## 6 -7.51429060
## 2 -1.37585603
## 4 -0.05981983
## 1  1.04454621
## 3  5.36077682
## 5  2.54464342
```

```
ranef(fm3)
```

```
## $Worker
##     MachineA    MachineB   MachineC
## 6 -5.5916012 -16.5838056 -5.0300736
## 2  0.1838843  -0.8033246 -4.2822824
## 4 -1.0238707   2.3284558 -1.4146284
```

```
## 1  0.3119890   2.5532237  0.9302963
## 3  6.9692219   7.7793499  4.4735106
## 5 -0.8496233   4.7261008  5.3231775
```

Note that model fm1, we can consider a single random effect term as

$$d_{ij} = b_i + b_{ij}$$

Predicted values for this new term d_{ij} given by

```
matrix(unlist(ranef(fm1)$`Machine:Worker`),6,3) + matrix(unlist(ranef(fm1)$Worker),6,3)
```

```
##            [,1]        [,2]       [,3]
## [1,] -5.5980169 -16.4900024 -5.0273291
## [2,]  0.1767089  -0.7690024 -4.3724886
## [3,] -1.0991915   2.3575464 -1.4742150
## [4,]  0.2943997   2.5445405  0.9303091
## [5,]  7.1383889   7.6601607  4.5458355
## [6,] -0.9122891   4.6967573  5.3978881
```

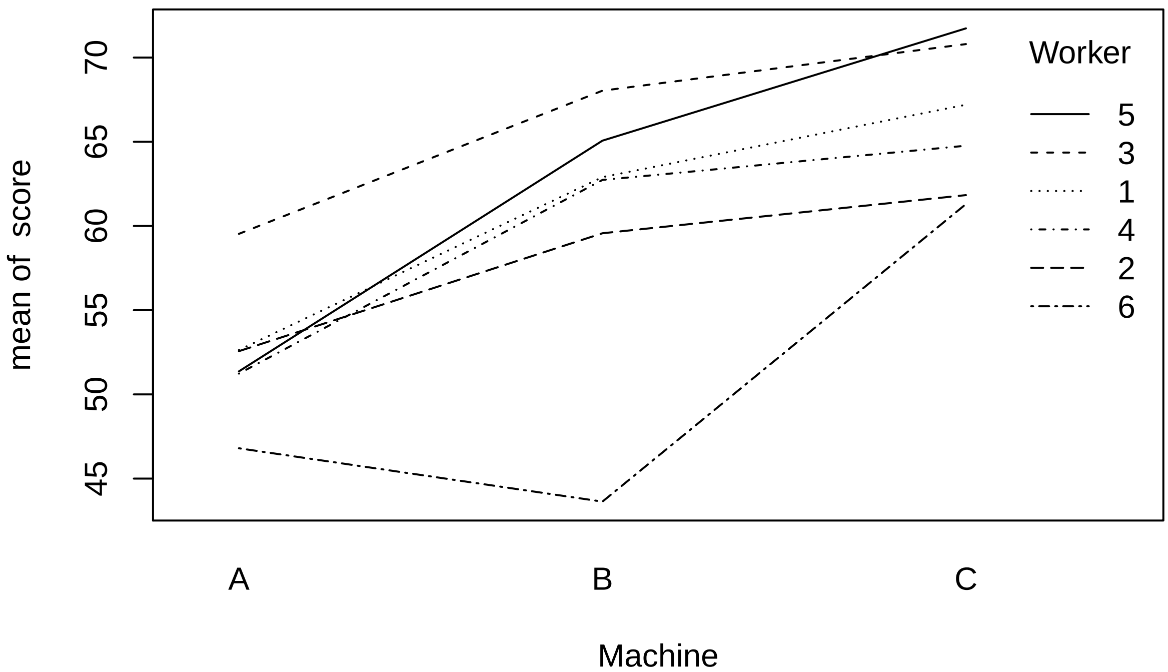
Compare these with the random effects in fm3:

```
ranef(fm3)
```

```
## $Worker
##     MachineA    MachineB   MachineC
## 6 -5.5916012 -16.5838056 -5.0300736
## 2  0.1838843  -0.8033246 -4.2822824
## 4 -1.0238707   2.3284558 -1.4146284
## 1  0.3119890   2.5532237  0.9302963
## 3  6.9692219   7.7793499  4.4735106
## 5 -0.8496233   4.7261008  5.3231775
```

Look again at interaction plot

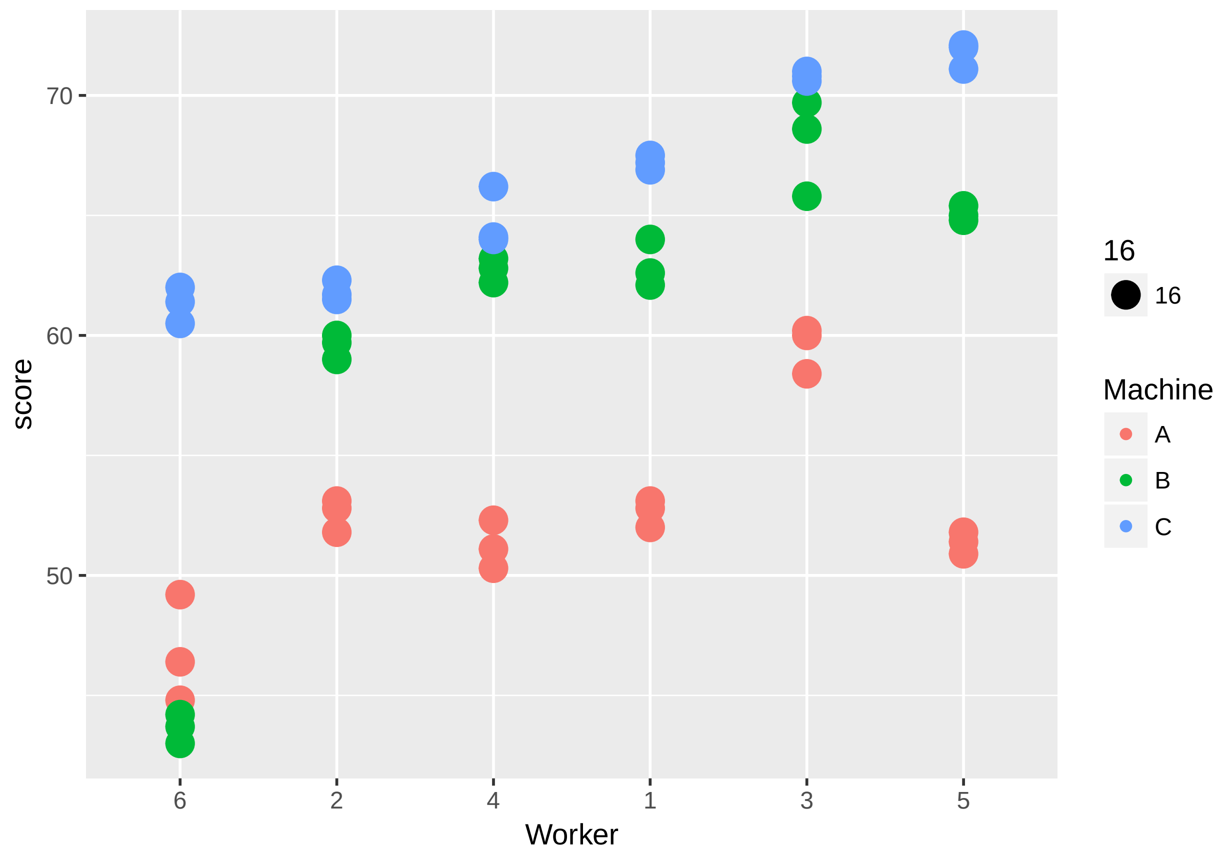
```
interaction.plot(Machine,Worker,score)
```



Can see why the MachineB variance is higher

Can also see why error variance is small

```
machinecolours<-rep(c("red","blue","green"),each=18)
#plot(as.numeric(Worker), score, col=machinecolours, pch=16)
qplot(Worker, score, col=Machine, cex=16)
```



Model fm1

$$Y_{ijk} = \beta_j + b_i + b_{ij} + \epsilon_{ijk}$$

with

$$b_i \sim N(0, \sigma_1^2),\ \ b_{ij} \sim N(0, \sigma_2^2),\ \ \epsilon_{ijk} \sim N(0, \sigma^2)$$

$$Var(Y_{ijk}) = \sigma_1^2 + \sigma_2^2 + \sigma^2 \ \text{ for all } \ i, j, k$$

Covariance between observations on the same worker on different machines is: σ_1^2, for any pair of machines.

Model fm3

$$Y_{ijk} = \beta_j + b_{ij} + \epsilon_{ijk}$$

with

$$b_{i1} \sim N(0, \sigma_1^2),\ \ b_{i2} \sim N(0, \sigma_2^2), b_{i3} \sim N(0, \sigma_3^2), \epsilon_{ijk} \sim N(0, \sigma^2)$$

so variance changes for each machine:

$$Var(Y_{i1k}) = \sigma_1^2 + \sigma^2$$

$$Var(Y_{i2k}) = \sigma_2^2 + \sigma^2$$
$$Var(Y_{i3k}) = \sigma_3^2 + \sigma^2$$

Covariance between observations on the same worker on different machines is allowed to vary for any pair of machines.

Model fm1 equivalent to fm3, with the constraints that variance is constant for each machine, and covariance of worker effects on any pair of machines is fixed across all pairs.

We will do formal testing later, but what would we look for in an exploratory data analysis to choose between the models?

Calculate means for each combination of worker and machine

```
cellmeans<-matrix(by(score,list(Worker,Machine),mean),6,3)
```

Note Worker is an ordered factor, ordered by increasing mean, so 1st row in cellmeans corresponds to worker 6, second row corresponds to worker 2 and so on.

Calculate variances of worker means for each machine

```
apply(cellmeans,2,var)
```

```
## [1] 16.94874 74.70385 19.57574
```

Variance for machine B is higher, suggesting unequal variance model fm3 might be suitable. Calculate correlations of worker means between each machine

```
cor(cellmeans)
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.7937749 0.6119436
## [2,] 0.7937749 1.0000000 0.7631595
## [3,] 0.6119436 0.7631595 1.0000000
```

unequal off-diagonal correlations would suggest trying fm3. However, sample size (6) for Workers is small, so not strong evidence to support fm3.

Note that these sample variances and correlations are very similar to the parameter estimates.

```
summary(fm3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ Machine - 1 + (Machine - 1 | Worker)
##    Data: Machines
##
## REML criterion at convergence: 208.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.39354 -0.51378  0.02691  0.47245  2.53339
##
## Random effects:
##  Groups   Name     Variance Std.Dev. Corr
##  Worker   MachineA 16.6405  4.0793
##           MachineB 74.3956  8.6253   0.80
##           MachineC 19.2675  4.3895   0.62 0.77
##  Residual           0.9246  0.9616
## Number of obs: 54, groups:  Worker, 6
##
```

```
## Fixed effects:
##          Estimate Std. Error t value
## MachineA   52.356      1.681   31.15
## MachineB   60.322      3.529   17.10
## MachineC   66.272      1.806   36.69
##
## Correlation of Fixed Effects:
##          MachnA MachnB
## MachineB 0.794
## MachineC 0.612  0.763
```