General Factor Design

Calvin Tran

The yield of a chemical process is being studied. The two factors of interest are temperature and pressure. Three levels of each factor are selected; however, only nine runs can be made in one day. The experimenter runs a complete replicate of the design on each day. The data are shown in the following table. Analyze the data, assuming that the days are blocks.

- Two factors, 2x3 design
- Blocking factor is day

In [14]:
```
library(ggplot2)

Yield = read.csv('/content/Yield.csv')
Yield$Blocks = as.factor(Yield$Blocks)
Yield$Temperature = as.factor(Yield$Temperature)
Yield$Pressure = as.factor(Yield$Pressure )

dim(Yield)
head(Yield)
```

18 · 4

A data.frame: 6 × 4

	Blocks	Temperature	Pressure	Yield
	<fct>	<fct>	<fct>	<dbl>
1	Day1	Low	250	86.3
2	Day1	Low	260	84.0
3	Day1	Low	270	85.8
4	Day1	Medium	250	88.5
5	Day1	Medium	260	87.3
6	Day1	Medium	270	89.0

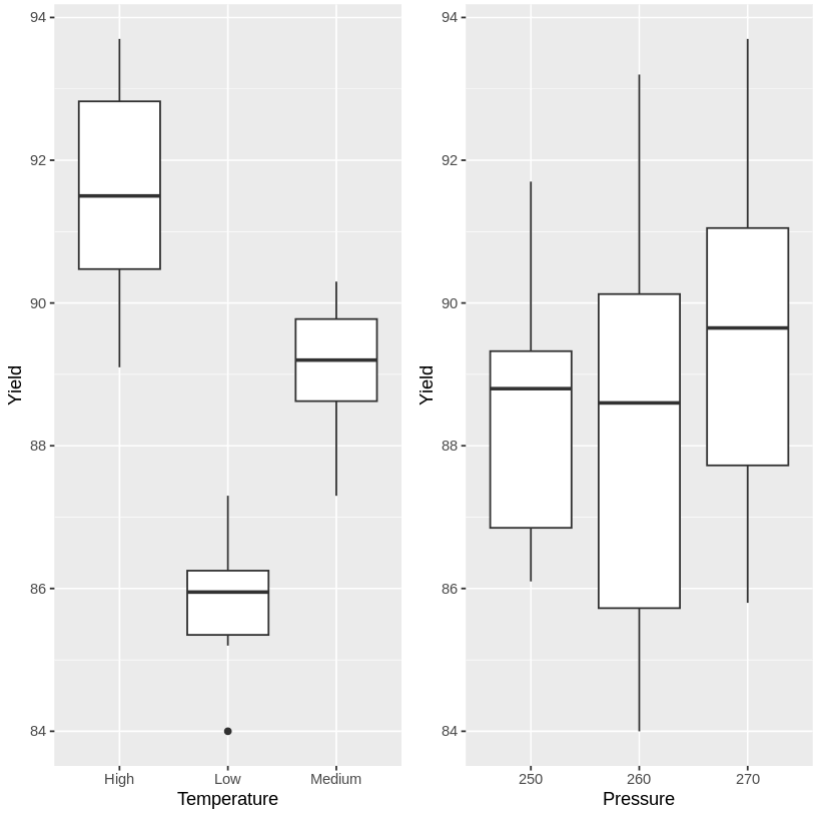
Examine Data Graphically

Main effects: Temperature and Pressure

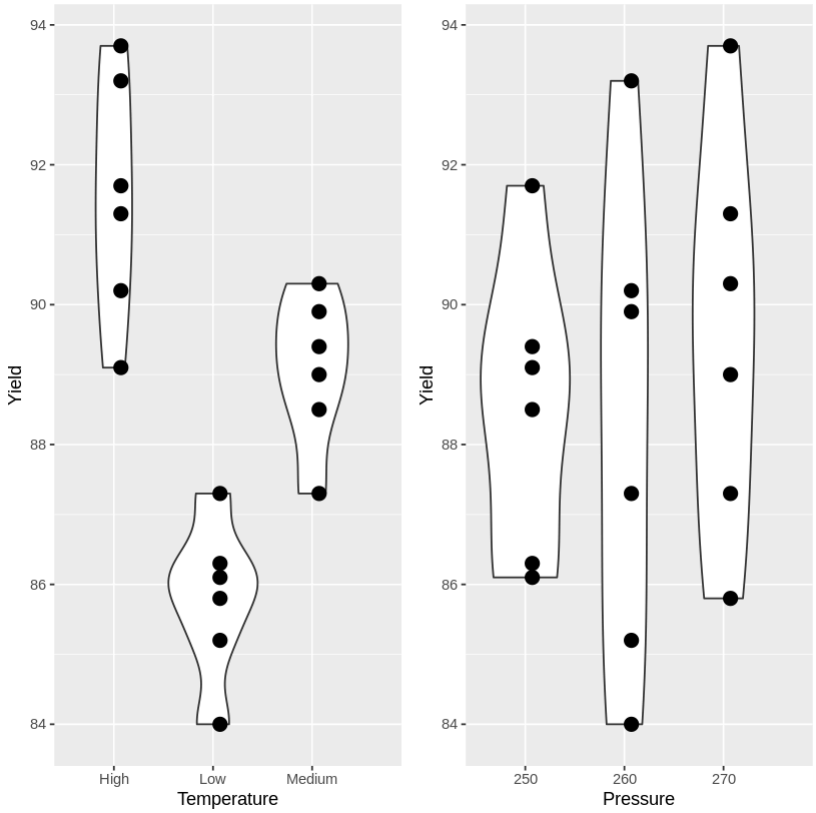
In [15]:
```
install.packages('patchwork',verbose=FALSE)
library(patchwork)
```

```
#boxplots
box1 = ggplot(Yield, aes(x=Temperature,y=Yield))+geom_boxplot()
box2 = ggplot(Yield, aes(x=Pressure,y=Yield))+geom_boxplot()
(box1|box2)
```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)
```



In [16]:
```
#violin plots
vio1 = ggplot(Yield, aes(group=Temperature, x=Temperature,y=Yield))+geom_violin()+geom
vio2 = ggplot(Yield, aes(group=Pressure, x=Pressure,y=Yield))+geom_violin()+geom_dotpl
vio1 | vio2
```



Plots

These plots suggest that the effect of temperature has a large impact on the resulting Yield.

The variance seen in pressure is likely explained as a result of being tested alongside different temperatures. The means for the different pressures appear roughly equal.

The interaction between factors will be explored later below.

Evaluate Necessity of Blocking Factor

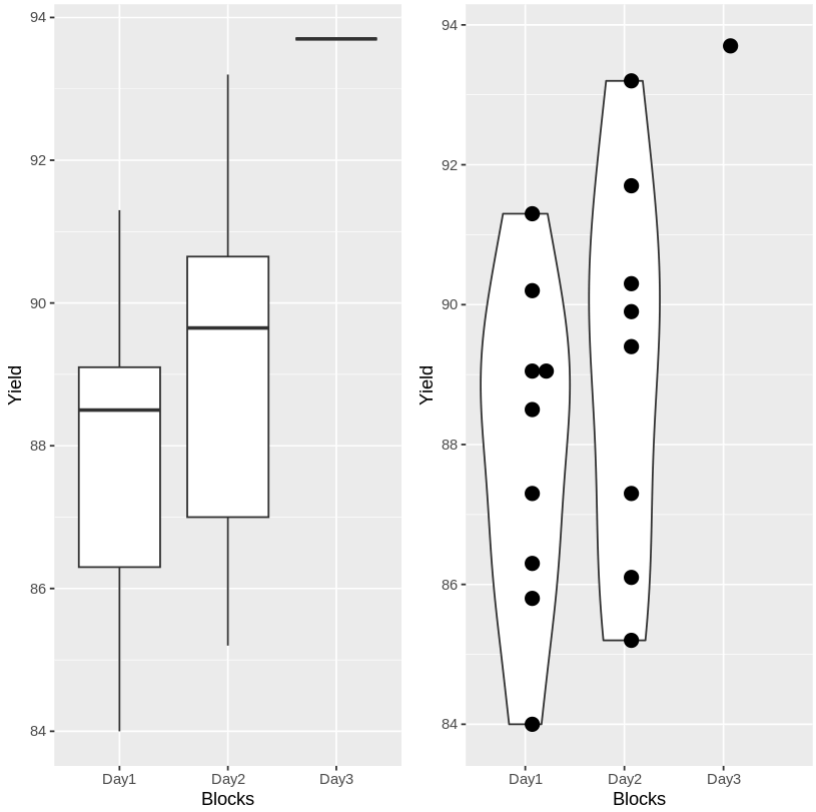
Compare the ratio of the Sum of Squares for the blocking factor to the Sum of Squares of the Overall Error

In [17]:
```
#examine the blocks graphically

box_block = ggplot(Yield, aes(x=Blocks,y=Yield))+geom_boxplot()
vio_block =ggplot(Yield, aes(group=Blocks, x=Blocks,y=Yield))+geom_violin()+geom_dotpl

box_block | vio_block
```

```
Warning message:
“Groups with fewer than two data points have been dropped.”
```



In [18]:
```
#Conduct analysis of variance to see the relative effect of blocking factor
Yield_aov = aov(Yield~Temperature*Pressure+Blocks, Yield)
summary(Yield_aov)
```

```
                    Df Sum Sq Mean Sq F value  Pr(>F)
Temperature          2  99.85   49.93  87.936 1.1e-05 ***
Pressure             2   5.51    2.75   4.850 0.04767 *
Blocks               2  13.45    6.72  11.844 0.00567 **
Temperature:Pressure 4   4.28    1.07   1.886 0.21777
Residuals            7   3.97    0.57
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Blocking factor

- Sum Sq of Blocks: 13.45
- Sum Sq of Residuals/Error: 3.97

Ratio: 13.45/3.97 = 3.3879

Graphically **the Yield means appear different** between the Day 1 and Day 2 blocks.

This indicates that the **blocking factor is worth keeping**. Blocking was helpful in improving the precision of the comparison of treatment means.

Conduct Analysis of Variance and test hypothesis on main effects

Conclusions using significance level 0.05

In [19]:
```
Yield_main_aov = aov(Yield~Temperature+Pressure+Blocks, Yield)
summary(Yield_main_aov)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Temperature  2  99.85   49.93  66.508 7.18e-07 ***
Pressure     2   5.51    2.75   3.668  0.06017 .
Blocks       2  13.45    6.72   8.958  0.00491 **
Residuals   11   8.26    0.75
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

F-statistic

```
# n = 2
# a = b = 3

# a-1 = 2
# (a*b)(n-1) = (3*3)(1) = 9
```

In [20]:
```
# Compute critical value of F
qf(c(0.05),2,9,lower.tail=FALSE)
```

4.25649472909375

In [21]:
```
#Multiple comparisons for the main effect of Temperature
install.packages('emmeans',verbose=FALSE)
library(emmeans)
print(emmeans(Yield_main_aov, pairwise ~ Temperature | Pressure))
```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)
```

```
$emmeans
Pressure = 250:
 Temperature emmean    SE df lower.CL upper.CL
 High          91.7 0.510 11     90.6     92.8
 Low           86.1 0.606 11     84.8     87.4
 Medium        89.4 0.606 11     88.0     90.7

Pressure = 260:
 Temperature emmean    SE df lower.CL upper.CL
 High          91.5 0.510 11     90.4     92.6
 Low           85.9 0.606 11     84.5     87.2
 Medium        89.2 0.606 11     87.8     90.5

Pressure = 270:
 Temperature emmean    SE df lower.CL upper.CL
 High          92.6 0.463 11     91.6     93.6
 Low           87.0 0.510 11     85.9     88.1
 Medium        90.3 0.510 11     89.2     91.4

Results are averaged over the levels of: Blocks
Confidence level used: 0.95

$contrasts
Pressure = 250:
 contrast      estimate    SE df t.ratio p.value
 High - Low        5.61 0.531 11  10.581  <.0001
 High - Medium     2.33 0.531 11   4.393  0.0028
 Low - Medium     -3.28 0.500 11  -6.564  0.0001

Pressure = 260:
 contrast      estimate    SE df t.ratio p.value
 High - Low        5.61 0.531 11  10.581  <.0001
 High - Medium     2.33 0.531 11   4.393  0.0028
 Low - Medium     -3.28 0.500 11  -6.564  0.0001

Pressure = 270:
 contrast      estimate    SE df t.ratio p.value
 High - Low        5.61 0.531 11  10.581  <.0001
 High - Medium     2.33 0.531 11   4.393  0.0028
 Low - Medium     -3.28 0.500 11  -6.564  0.0001

Results are averaged over the levels of: Blocks
P value adjustment: tukey method for comparing a family of 3 estimates
```

These estimates indicate that as temperature increases from Low to Medium to High, the mean yield also increases. This is consistent for all pressures.

In [22]:
```
print(emmeans(Yield_main_aov, pairwise ~ Pressure | Temperature)$contrasts)
```

```
Temperature = High:
 contrast                  estimate    SE df t.ratio p.value
 Pressure250 - Pressure260    0.217 0.500 11   0.433  0.9026
 Pressure250 - Pressure270   -0.914 0.531 11  -1.722  0.2405
 Pressure260 - Pressure270   -1.131 0.531 11  -2.131  0.1287

Temperature = Low:
 contrast                  estimate    SE df t.ratio p.value
 Pressure250 - Pressure260    0.217 0.500 11   0.433  0.9026
 Pressure250 - Pressure270   -0.914 0.531 11  -1.722  0.2405
 Pressure260 - Pressure270   -1.131 0.531 11  -2.131  0.1287

Temperature = Medium:
 contrast                  estimate    SE df t.ratio p.value
 Pressure250 - Pressure260    0.217 0.500 11   0.433  0.9026
 Pressure250 - Pressure270   -0.914 0.531 11  -1.722  0.2405
 Pressure260 - Pressure270   -1.131 0.531 11  -2.131  0.1287

Results are averaged over the levels of: Blocks
P value adjustment: tukey method for comparing a family of 3 estimates
```

On the other hand, these p-values for these estimates indicate that changing the pressure has no significant effect on the mean.

Main Effects

Using a 0.05 significance level, it appears that **the temperature factor affects the mean** while the **pressure factor does not**.

Critical F-value = 4.256

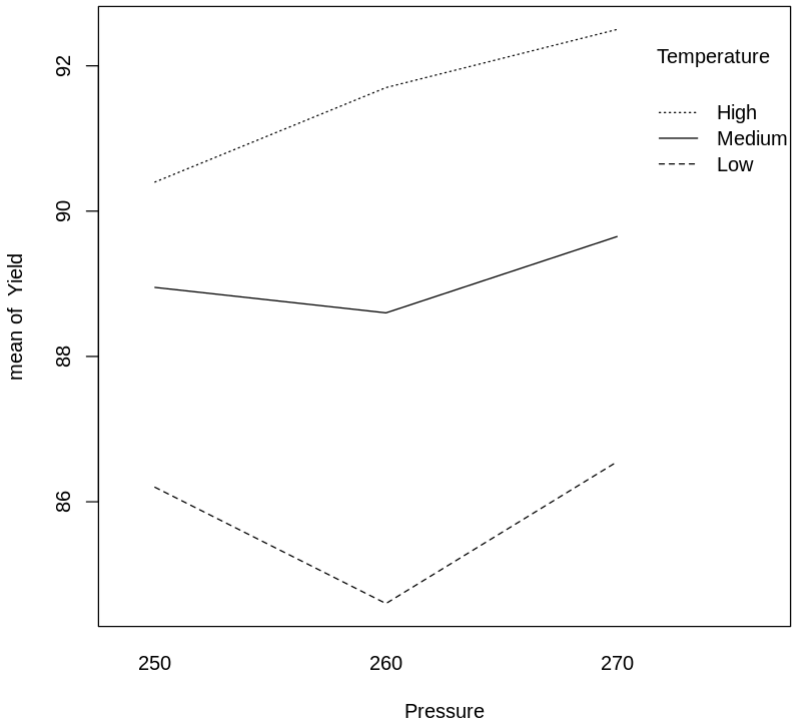
- The F-value for temperature, at 66.508 is larger than this critical value. Along with a very small p-value well below the significance level, This indicates a significant difference in means.
- The F-value for pressure, at 3.668 is smaller than this critical value. Along with a p-value higher than the significance level, This indicates no difference in means, failing to reject the null hypothesis.

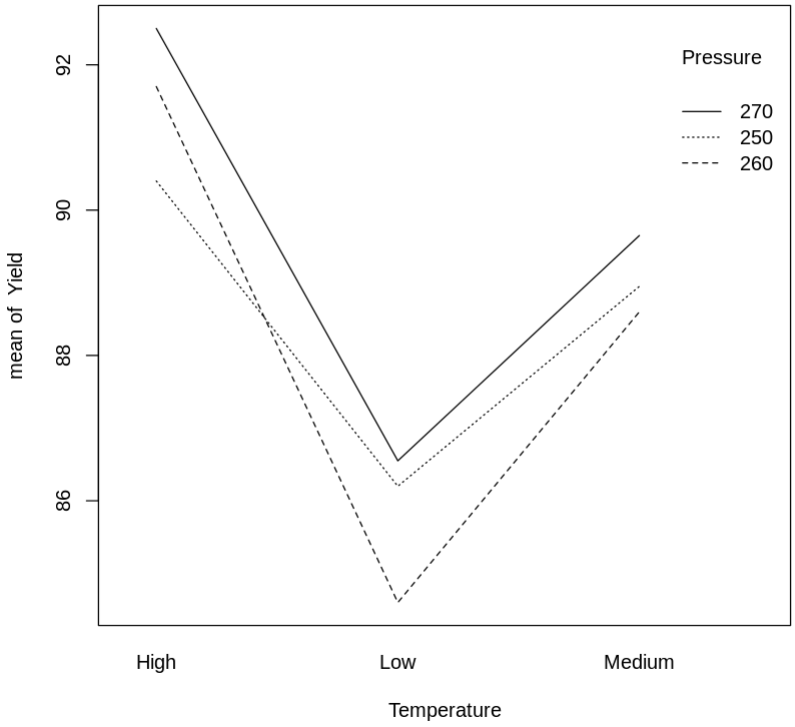
Construct Interaction Plot & test on interaction of factors

Conclusions using significance level 0.05

In [23]:
```
#interaction plot
interaction1 = with(Yield, interaction.plot(Pressure,Temperature,Yield))
interaction2 = with(Yield, interaction.plot(Temperature,Pressure,Yield))

(interaction1|interaction2)
```





While two of the lines seem rather parallel, in the first chart the line for High temperature differs for pressure=250, and in the second chart the lines for pressure=250 and pressure=260 intersect.

This indicates that there is an interaction between factors, and this seems to exist at high temperatures at 250 pressure.

In [24]:
```
#aov
Yield_aov = aov(Yield~Temperature*Pressure+Blocks, Yield)
summary(Yield_aov)
```

```
                    Df Sum Sq Mean Sq F value  Pr(>F)
Temperature          2  99.85   49.93  87.936 1.1e-05 ***
Pressure             2   5.51    2.75   4.850 0.04767 *
Blocks               2  13.45    6.72  11.844 0.00567 **
Temperature:Pressure 4   4.28    1.07   1.886 0.21777
Residuals            7   3.97    0.57
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In [25]:
```
print(emmeans(Yield_aov, pairwise ~ Temperature | Pressure)$contrasts)
```

```
Pressure = 250:
 contrast      estimate    SE df t.ratio p.value
 High - Low        4.20 0.754  7   5.574  0.0021
 High - Medium     1.45 0.754  7   1.924  0.2018
 Low - Medium     -2.75 0.754  7  -3.650  0.0196

Pressure = 260:
 contrast      estimate    SE df t.ratio p.value
 High - Low        7.10 0.754  7   9.423  0.0001
 High - Medium     3.10 0.754  7   4.114  0.0109
 Low - Medium     -4.00 0.754  7  -5.309  0.0028

Pressure = 270:
 contrast      estimate    SE df t.ratio p.value
 High - Low        5.56 0.942  7   5.899  0.0015
 High - Medium     2.46 0.942  7   2.608  0.0794
 Low - Medium     -3.10 0.754  7  -4.114  0.0109

Results are averaged over the levels of: Blocks
P value adjustment: tukey method for comparing a family of 3 estimates
```

These results differ slightly from the model without the interaction factor.

At pressure = 250 and at pressure = 270, the contrasts between high and medium temperature is higher than the significance level.

This indicates that **at certain pressures, increasing the temperature from "medium" to "high" has no effect.** The resulting mean yield is the same.

In [26]:
```
print(emmeans(Yield_aov, pairwise ~ Pressure | Temperature)$contrasts)
```

```
Temperature = High:
 contrast                  estimate    SE df t.ratio p.value
 Pressure250 - Pressure260   -1.300 0.754  7  -1.725  0.2619
 Pressure250 - Pressure270   -1.706 0.942  7  -1.812  0.2342
 Pressure260 - Pressure270   -0.406 0.942  7  -0.431  0.9039

Temperature = Low:
 contrast                  estimate    SE df t.ratio p.value
 Pressure250 - Pressure260    1.600 0.754  7   2.123  0.1544
 Pressure250 - Pressure270   -0.350 0.754  7  -0.464  0.8897
 Pressure260 - Pressure270   -1.950 0.754  7  -2.588  0.0816

Temperature = Medium:
 contrast                  estimate    SE df t.ratio p.value
 Pressure250 - Pressure260    0.350 0.754  7   0.464  0.8897
 Pressure250 - Pressure270   -0.700 0.754  7  -0.929  0.6406
 Pressure260 - Pressure270   -1.050 0.754  7  -1.393  0.3942

Results are averaged over the levels of: Blocks
P value adjustment: tukey method for comparing a family of 3 estimates
```

These p-values still indicate that changing the pressure has no effect on the yield.