Case Study 7

Kamryn Benusa

Part 1: Agricultural Field Trial

Split plot designs are commonly used in agricultural studies. In one agricultural field trial, the objective was to examine the effects of two crop varieties and four different irrigation methods on yield. Eight fields were available for the study but only one type of irrigation can be used in a field which is randomly assigned. The fields are divided into two parts with a different variety planted in each half (which is randomly assigned). Yield is measured for each half-plot.

```
#read in data
irrigation1 <- read_csv('irrigation.csv')
```

```
## Rows: 16 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): field, irrigation, variety
## dbl (1): yield
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
irrigation1
```

```
## # A tibble: 16 x 4
##    field irrigation variety yield
##    <chr> <chr>      <chr>   <dbl>
##  1 f1    i1         v1       35.4
##  2 f1    i1         v2       37.9
##  3 f2    i2         v1       36.7
##  4 f2    i2         v2       38.2
##  5 f3    i3         v1       34.8
##  6 f3    i3         v2       36.4
##  7 f4    i4         v1       39.5
##  8 f4    i4         v2       40
##  9 f5    i1         v1       41.6
## 10 f5    i1         v2       40.3
## 11 f6    i2         v1       42.7
## 12 f6    i2         v2       41.6
## 13 f7    i3         v1       43.6
## 14 f7    i3         v2       42.8
## 15 f8    i4         v1       44.5
## 16 f8    i4         v2       47.6
```

```r
irrigation1 <- irrigation1 %>%
  mutate(field = as_factor(field),
         irrigation = as_factor(irrigation),
         variety = as_factor(variety))
irrigation1
```

```
## # A tibble: 16 x 4
##    field irrigation variety yield
##    <fct> <fct>      <fct>   <dbl>
##  1 f1    i1         v1       35.4
##  2 f1    i1         v2       37.9
##  3 f2    i2         v1       36.7
##  4 f2    i2         v2       38.2
##  5 f3    i3         v1       34.8
##  6 f3    i3         v2       36.4
##  7 f4    i4         v1       39.5
##  8 f4    i4         v2       40
##  9 f5    i1         v1       41.6
## 10 f5    i1         v2       40.3
## 11 f6    i2         v1       42.7
## 12 f6    i2         v2       41.6
## 13 f7    i3         v1       43.6
## 14 f7    i3         v2       42.8
## 15 f8    i4         v1       44.5
## 16 f8    i4         v2       47.6
```

1. [2pts] Identify the following:

a. Whole plot eu

Answer: The field

b. Whole plot treatment

Answer: Irrigation

c. Split plot eu

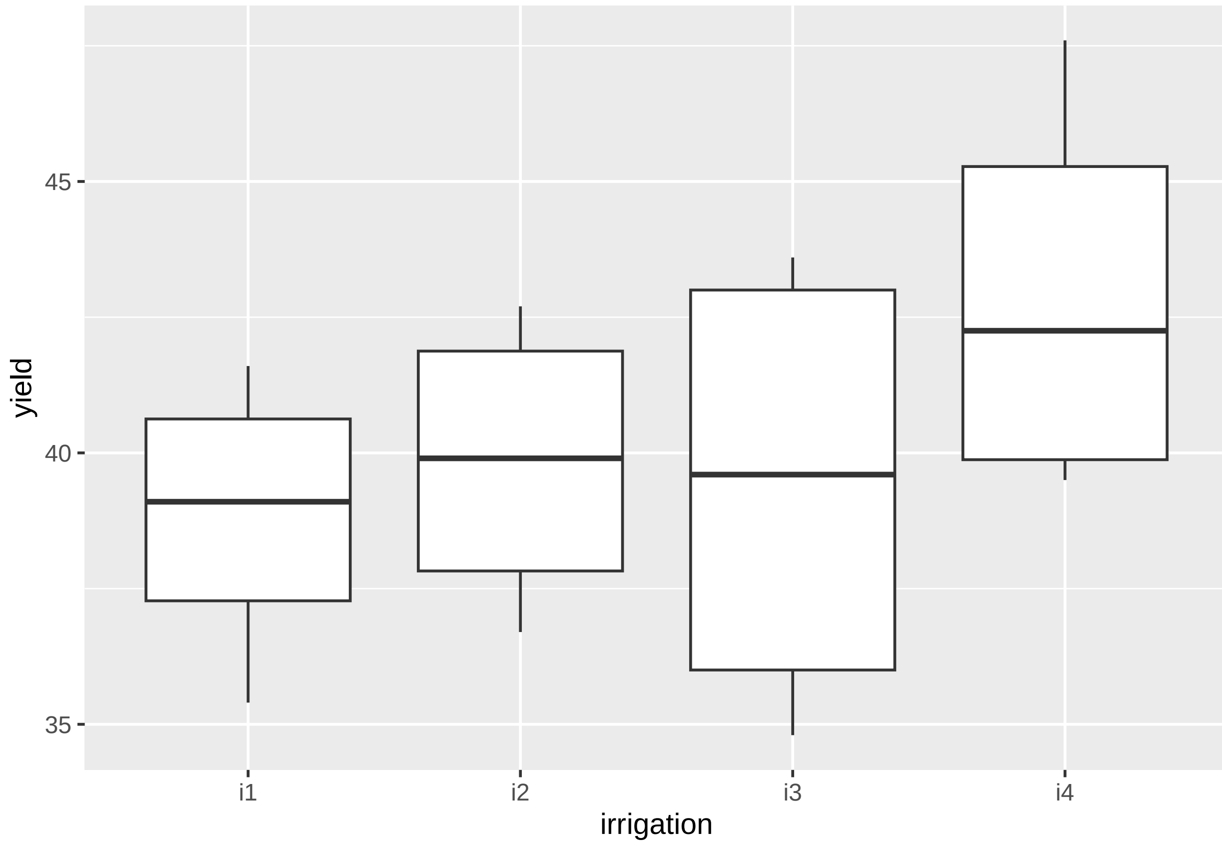
Answer: Half a field

d. Split plot treatment

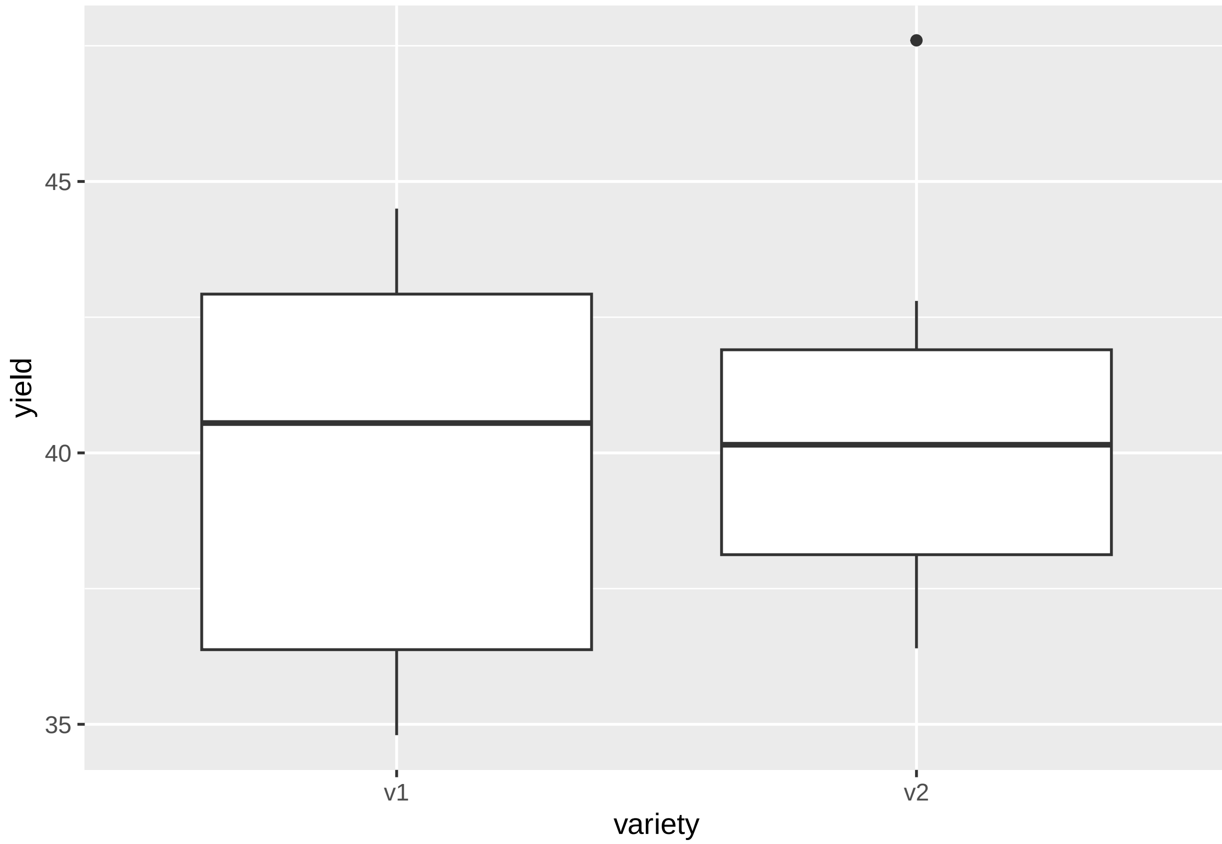
Answer: Crop variety

2. [4pts] Using the appropriate graphical display, describe how irrigation method and variety affect yield.

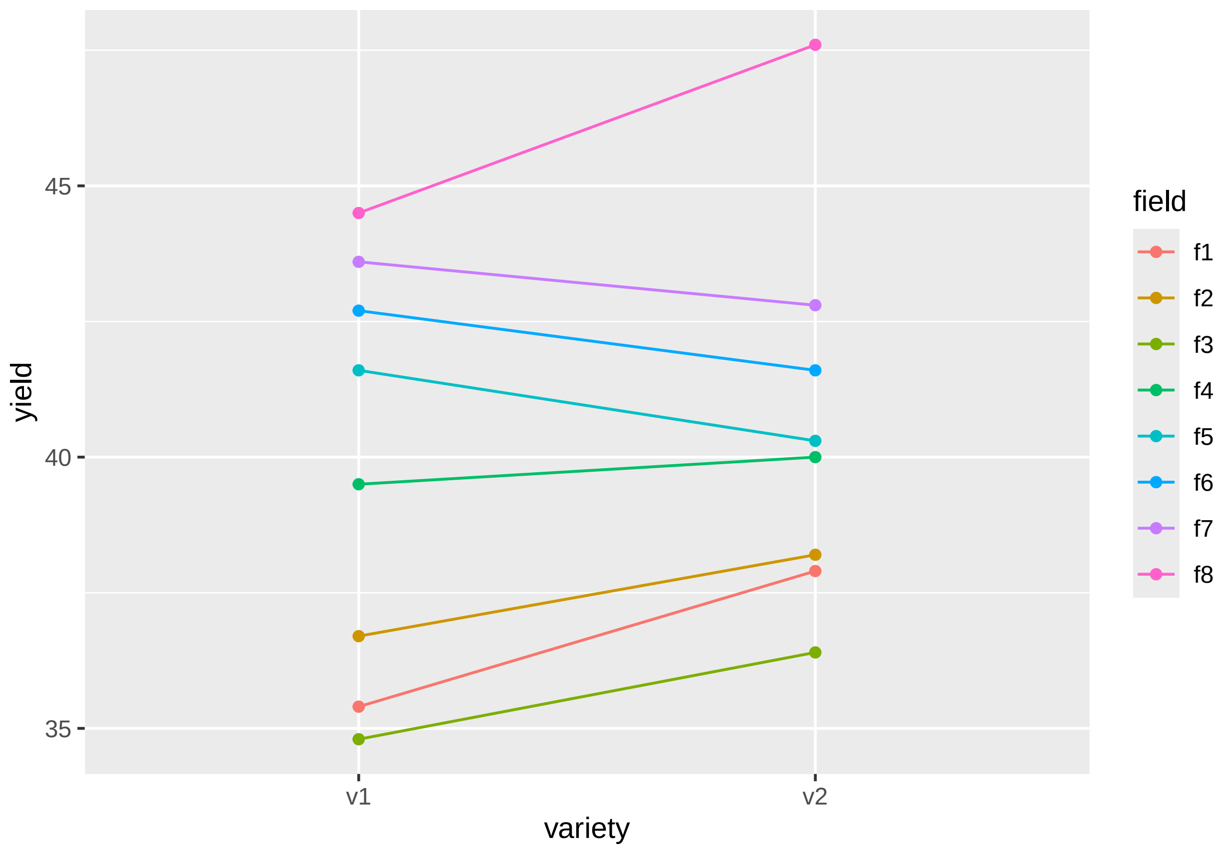
```r
#whole plot-treatment effect (irrigation)
irrigation1 %>% ggplot(aes(x = irrigation, y = yield)) +
  geom_boxplot()
```




```
#split plot treatment effect (crop variety)
irrigation1 %>% ggplot(aes(x = variety, y = yield)) +
  geom_boxplot()
```



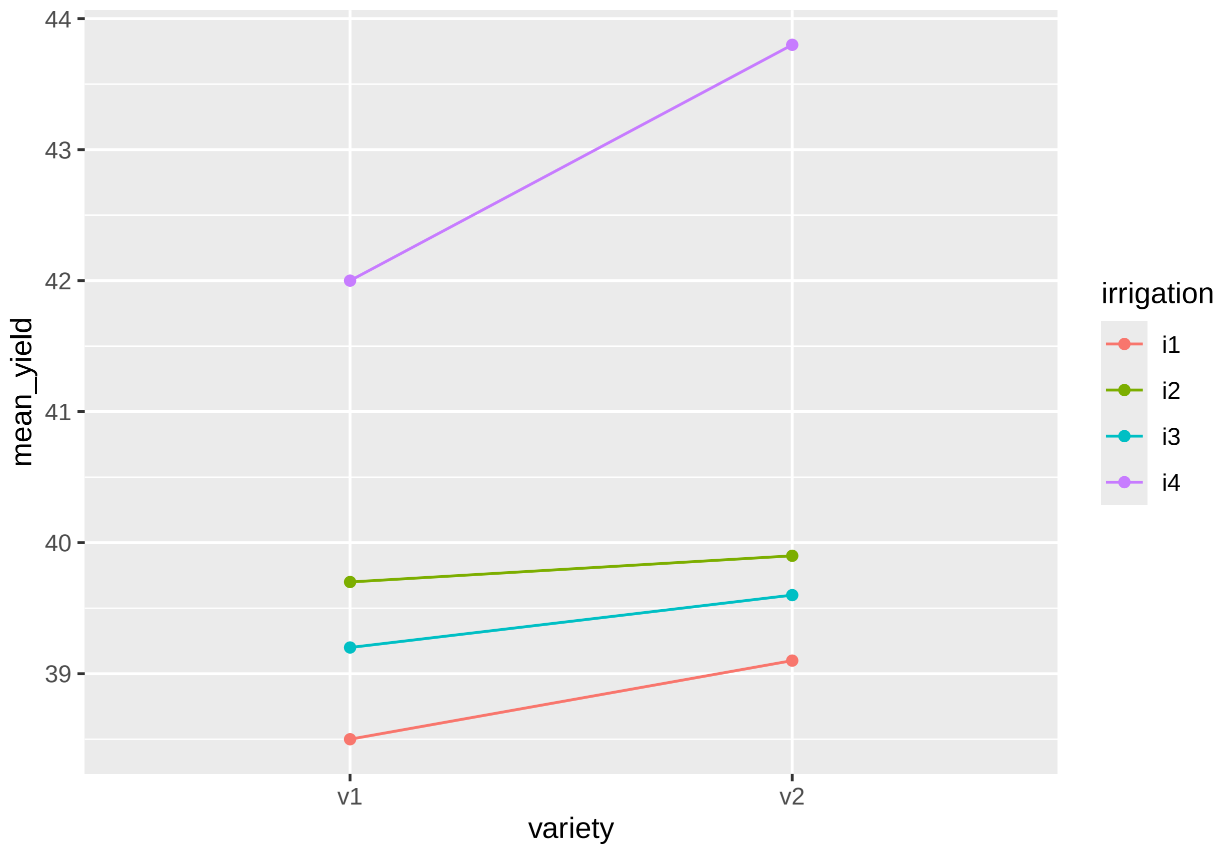

```
# variety and field
irrigation1 %>% ggplot(aes(x = variety, y = yield)) +
  geom_line(aes(group = field, color = field)) +
  geom_point(aes(color = field))
```




```
# irrigation and field
irrigation_mean <- irrigation1 %>%
  group_by(irrigation, variety) %>%
  summarize(mean_yield = mean(yield))
```

```
## `summarise()` has grouped output by 'irrigation'. You can override using the
## `.groups` argument.
```

```
irrigation_mean %>% ggplot(aes(x = variety, y = mean_yield))+
  geom_line(aes(group = irrigation, color = irrigation)) +
  geom_point(aes(color = irrigation))
```



Answer: For irrigation, it appears that irrigation 4 has a higher yield on average than irrigations 1, 2, and 3. Irrigations 1, 2, and 3 all have similar average yields. The distribution of yield for irrigations 1, 2, and 3 are roughly symmetric, while the distribution for irrigation 4 appears slightly right-skewed. The spread for irrigation 3 appears to be greater than the spread for irrigations 1, 2, and 4. There are no potential outliers.

For crop variety, the distributions of yield for both variety 1 and 2 are roughly symmetric. The average yield is similar for both variety 1 and 2. The spread of yield for variety 1 is greater than the spread for variety 2. There is one potential outlier at the maximum yield for crop variety 2.

There does not appear to be an interaction between field and variety, or irrigation and variety.

3. [4pts] Let's assume that all assumptions are reasonably satisfied for this analysis (normality is a bit questionable if you care to look). Using the appropriate models and statistics, explain which treatment factors affect yield. In your explanation make sure to include appropriate statistics, df, and p-values.

```
irrig_reml <- lmer(yield ~ variety * irrigation + (1|field), data = irrigation1)
anova(irrig_reml)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                    Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## variety            2.2500 2.25000     1     4  1.0676 0.3599
## irrigation         2.4545 0.81818     3     4  0.3882 0.7685
## variety:irrigation 1.5500 0.51667     3     4  0.2452 0.8612
```

```
# remove interaction
irrig_reml2 <- lmer(yield ~ variety + irrigation + (1|field), data = irrigation1)
anova(irrig_reml2)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##            Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## variety    2.2500  2.2500     1     7  1.5782 0.2493
## irrigation 1.6605  0.5535     3     4  0.3882 0.7685
```

```r
ranova(irrig_reml2)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## yield ~ variety + irrigation + (1 | field)
##             npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>         7 -27.398 68.796
## (1 | field)    6 -33.371 78.741 11.946  1  0.0005477 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: Variety does not affect yield (F: 1.58, df: 1 and 7, p-value: 0.25) and irrigation does not affect yield (F: 0.39, df: 3 and 4, p-value: 0.77). There is strong evidence that field affects yield (χ^2 : 11.95 and p-value: 0.0006).

4. [2pts] In the results you used in Question 3 you hopefully noticed that the denominator degrees of freedom differ for the tests of the main effect of irrigation and variety. Explain why this occurs.

Answer: The denominator degrees of freedom differ for the tests of the main effect of irrigation and variety because the calculation for the denominator degrees of freedom for irrigation uses the number of whole plot eu's, while the calculation for the denominator degrees of freedom for variety uses the split plot eu's.

Part 2: Glue Strength

An assembly company is testing four different formulations of an industrial glue to examine which is most effective. The glue is applied to join parts and the tensile strength (in pounds) is measured after application. Each glue is used in five applications. While the same amount of glue is to be applied, there are differences among the thickness of the application of the glue (in 0.01 inches) which is known to impact strength.

```r
#read in data
glue <- read_csv('glue.csv')
```

```
## Rows: 20 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (3): Strength, Glue, Thickness
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
glue
```

```
## # A tibble: 20 x 3
##    Strength  Glue Thickness
##       <dbl> <dbl>     <dbl>
##  1       47     1        13
##  2       45     1        14
##  3       49     1        12
##  4       47     1        12
##  5       45     1        14
##  6       49     2        12
##  7       50     2        10
##  8       49     2        11
##  9       48     2        12
## 10       47     2        14
## 11       46     3        15
## 12       46     3        14
## 13       48     3        11
## 14       48     3        11
## 15       48     3        10
## 16       45     4        16
## 17       45     4        15
## 18       52     4        10
## 19       48     4        12
## 20       50     4        11
```

```r
#convert to factor as needed
glue <- glue %>%
  mutate(Glue = as_factor(Glue))
glue
```

```
## # A tibble: 20 x 3
##    Strength Glue  Thickness
##       <dbl> <fct>     <dbl>
##  1       47 1            13
##  2       45 1            14
##  3       49 1            12
##  4       47 1            12
##  5       45 1            14
##  6       49 2            12
##  7       50 2            10
##  8       49 2            11
##  9       48 2            12
## 10       47 2            14
## 11       46 3            15
## 12       46 3            14
## 13       48 3            11
## 14       48 3            11
## 15       48 3            10
## 16       45 4            16
## 17       45 4            15
## 18       52 4            10
## 19       48 4            12
## 20       50 4            11
```

5. [2pts] One way to design this study would be to use the thickness of the glue application as a blocking

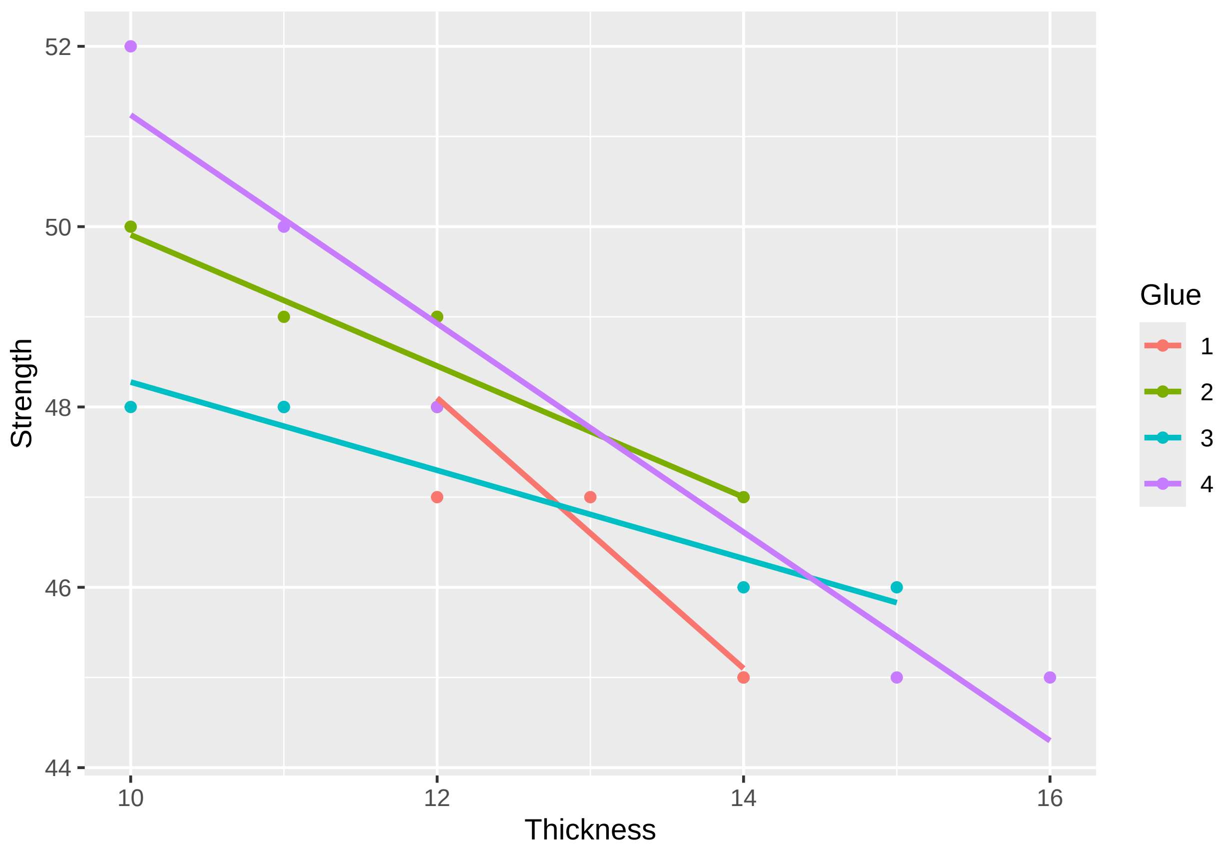
variable. Explain the advantage(s) of including thickness as a covariate instead of as a blocking variable.

Answer: Including thickness as a covariate instead of as a blocking variable allows for analysis of the effect of thickness. With blocking, the effect of the blocks is not examined. Additionally, including thickness as a covariate accounts for the fact that we cannot randomly assign thickness.

6. [4pts] Obtain the appropriate plot that examines the relationship between thickness and glue formulation with strength. Explain what this plot suggests.

```
glue %>% ggplot(aes(x = Thickness, y = Strength)) + # explanatory on x axis, response on y axis
  geom_point(aes(color = Glue)) + # adds color to point
  geom_smooth(method = "lm", se = F, aes(group = Glue, color = Glue))
```

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



Answer: This graph suggests that as thickness increases, the strength of the glue decreases. Additionally, the formulations of glue appear to have a similar strength. There could be a possible interaction as the slopes of the lines differ.

7. [3pts] In class we discussed several potential models of interest (e.g. parallel slopes). Using the appropriate tools, explain which model best describes the data. Note: the assumptions for this analysis are reasonably satisfied (feel free to check!).

```r
glue_int <- lm(Strength ~ Glue * Thickness, data = glue)
anova(glue_int)
```

```
## Analysis of Variance Table
##
## Response: Strength
##                Df Sum Sq Mean Sq  F value    Pr(>F)
## Glue            3 11.600   3.867   8.9487  0.002184 **
## Thickness       1 47.376  47.376 109.6431 2.173e-07 ***
## Glue:Thickness  3  6.639   2.213   5.1215  0.016462 *
## Residuals      12  5.185   0.432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: The separate lines model best describes the data since the effect of formulation of glue depends on thickness of application (F: 5.12, df: 3, p-value: 0.017).

8. [2pts] Report the estimated model for glue formulation 3.

```r
summary(glue_int)
```

```
##
## Call:
## lm(formula = Strength ~ Glue * Thickness, data = glue)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10000 -0.28723 -0.04104  0.25957  0.90000
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     66.1000     4.2828  15.434  2.8e-09 ***
## Glue2           -8.9182     5.0265  -1.774  0.10138
## Glue3          -12.9298     4.6744  -2.766  0.01708 *
## Glue4           -3.2940     4.5902  -0.718  0.48673
## Thickness       -1.5000     0.3287  -4.564  0.00065 ***
## Glue2:Thickness  0.7727     0.3964   1.949  0.07500 .
## Glue3:Thickness  1.0106     0.3619   2.792  0.01628 *
## Glue4:Thickness  0.3433     0.3523   0.974  0.34915
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6573 on 12 degrees of freedom
## Multiple R-squared:  0.9268, Adjusted R-squared:  0.884
## F-statistic: 21.69 on 7 and 12 DF,  p-value: 6.543e-06
```

Answer: 66.1 - 12.9298 - (1.5 + 1.0106)Thickness

9. [5pts] Using the appropriate tools, describe how the four glue formulations compare in terms of strength.

```r
means_Q1<-emmeans(glue_int,~Glue*Thickness,at=list(Thickness=quantile(glue$Thickness,0.25)))
means_median<-emmeans(glue_int,~Glue*Thickness,at=list(Thickness=median(glue$Thickness)))
means_Q3<-emmeans(glue_int,~Glue*Thickness,at=list(Thickness=quantile(glue$Thickness,0.75)))
confint(pairs(means_Q1))
```

```
##  contrast                              estimate    SE df lower.CL upper.CL
##  Glue1 Thickness11 - Glue2 Thickness11    0.418 0.798 12  -1.9502    2.787
##  Glue1 Thickness11 - Glue3 Thickness11    1.813 0.799 12  -0.5587    4.184
##  Glue1 Thickness11 - Glue4 Thickness11   -0.482 0.811 12  -2.8889    1.925
##  Glue2 Thickness11 - Glue3 Thickness11    1.395 0.487 12  -0.0518    2.841
##  Glue2 Thickness11 - Glue4 Thickness11   -0.900 0.506 12  -2.4039    0.603
##  Glue3 Thickness11 - Glue4 Thickness11   -2.295 0.508 12  -3.8034   -0.786
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 4 estimates
```

```r
confint(pairs(means_median))
```

```
##  contrast                              estimate    SE df lower.CL upper.CL
##  Glue1 Thickness12 - Glue2 Thickness12   -0.355 0.532 12  -1.9334    1.224
##  Glue1 Thickness12 - Glue3 Thickness12    0.802 0.531 12  -0.7739    2.378
##  Glue1 Thickness12 - Glue4 Thickness12   -0.825 0.540 12  -2.4274    0.777
##  Glue2 Thickness12 - Glue3 Thickness12    1.157 0.419 12  -0.0879    2.401
##  Glue2 Thickness12 - Glue4 Thickness12   -0.471 0.430 12  -1.7482    0.807
##  Glue3 Thickness12 - Glue4 Thickness12   -1.628 0.429 12  -2.9013   -0.354
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 4 estimates
```

```r
confint(pairs(means_Q3))
```

```
##  contrast                              estimate    SE df lower.CL upper.CL
##  Glue1 Thickness14 - Glue2 Thickness14   -1.900 0.720 12    -4.04    0.238
##  Glue1 Thickness14 - Glue3 Thickness14   -1.219 0.596 12    -2.99    0.551
##  Glue1 Thickness14 - Glue4 Thickness14   -1.512 0.551 12    -3.15    0.125
##  Glue2 Thickness14 - Glue3 Thickness14    0.681 0.696 12    -1.39    2.748
##  Glue2 Thickness14 - Glue4 Thickness14    0.388 0.659 12    -1.57    2.343
##  Glue3 Thickness14 - Glue4 Thickness14   -0.293 0.520 12    -1.84    1.251
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 4 estimates
```

Answer: At thicknesses 11 and 12, glue 4 is stronger than glue 3, while at thickness 14 they are similar. All other glue types at each thickness examined are similar.