Lab 5

Adam Shen

October 14, 2020

Packages

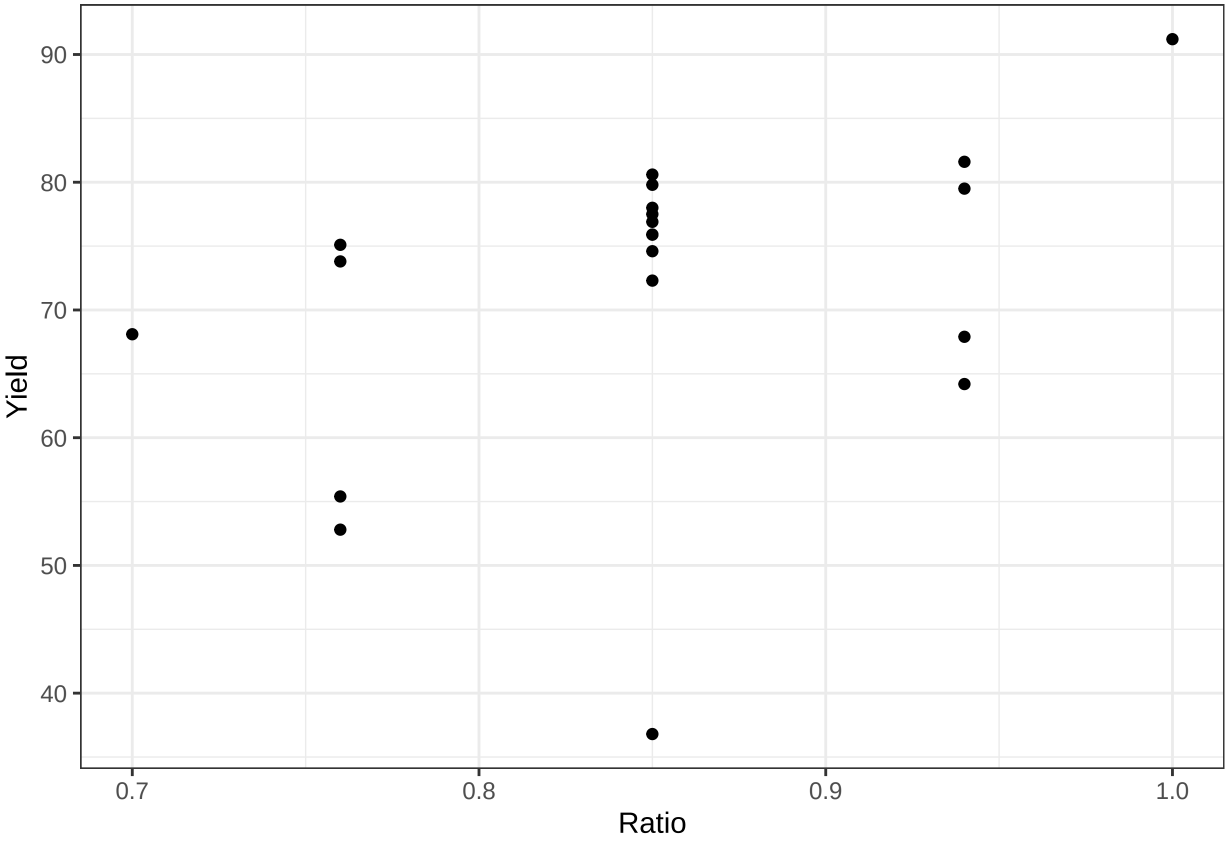
```
library(car)
library(ggplot2)
theme_set(theme_bw())
```

Chemical production data

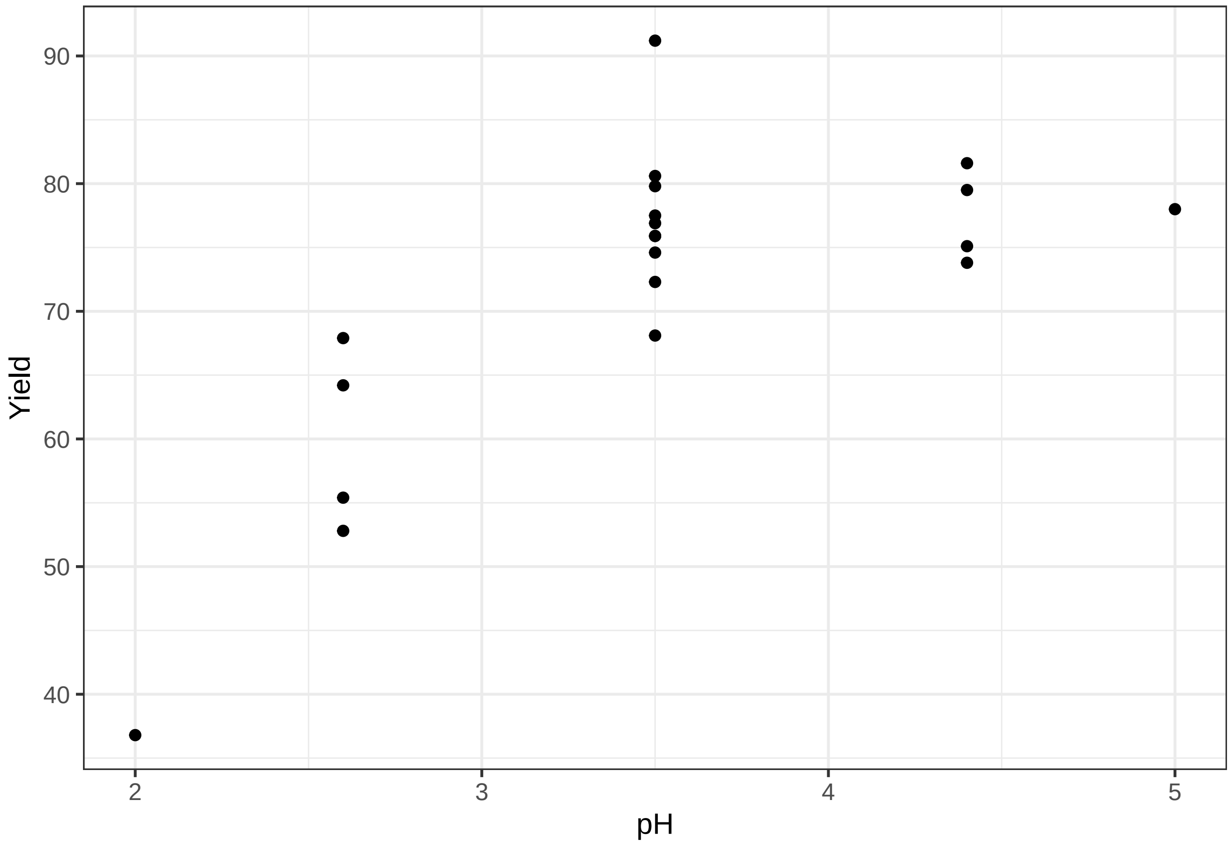
```
chemical <- read.table("./chemprod.txt", header=TRUE)
```

Scatterplots

```
ggplot(chemical, aes(x=Ratio, y=Yield))+
  geom_point()
```



```
ggplot(chemical, aes(x=pH, y=Yield))+
  geom_point()
```



Data Prep

```
chemical2 <- transform(chemical,
                       CRatio = Ratio - mean(Ratio),
                       CpH = pH - mean(pH))
```

Fit multiple linear regression model

```
model <- lm(Yield ~ CRatio + CpH + (CRatio*CpH) + I(CRatio^2) + I(CpH^2), x=TRUE, data=chemical2)
# Putting brackets around the interaction term makes it easier to read when we do ANOVA
summary(model)
```

```
##
## Call:
## lm(formula = Yield ~ CRatio + CpH + (CRatio * CpH) + I(CRatio^2) +
##     I(CpH^2), data = chemical2, x = TRUE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4921 -2.1355 -0.5197  1.7556  4.9649
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  76.4197     1.1053  69.142  < 2e-16 ***
## CRatio       61.1475     9.6141   6.360 1.77e-05 ***
```

```
## CpH            11.3415     0.9614  11.797 1.16e-08 ***
## I(CRatio^2)    82.3584   104.5722   0.788    0.444    
## I(CpH^2)       -9.0653     1.0457  -8.669 5.31e-07 ***
## CRatio:CpH    -18.0556    13.9053  -1.298    0.215    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.186 on 14 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9303 
## F-statistic: 51.75 on 5 and 14 DF,  p-value: 1.56e-08
```

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Yield
##              Df  Sum Sq Mean Sq  F value    Pr(>F)    
## CRatio        1  410.54  410.54  40.4522 1.768e-05 ***
## CpH           1 1412.36 1412.36 139.1640 1.164e-08 ***
## I(CRatio^2)   1   23.31   23.31   2.2972    0.1519    
## I(CpH^2)      1  762.69  762.69  75.1506 5.313e-07 ***
## CRatio:CpH    1   17.11   17.11   1.6860    0.2151    
## Residuals    14  142.08   10.15                       
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obtaining the variance-covariance matrix

```
res <- resid(model)
MSE <- sum(res^2)/model$df.residual
X <- model$x
XTXinv <- solve(t(X) %*% X)
```

```
(Vhatb <- MSE*XTXinv)
```

```
##               (Intercept)        CRatio           CpH   I(CRatio^2)
## (Intercept)  1.221590e+00  0.000000e+00  1.476340e-16 -6.504059e+01
## CRatio       0.000000e+00  9.243070e+01  0.000000e+00  0.000000e+00
## CpH          1.476340e-16  0.000000e+00  9.243070e-01 -6.986623e-15
## I(CRatio^2) -6.504059e+01  0.000000e+00 -6.986623e-15  1.093535e+04
## I(CpH^2)    -6.504059e-01  0.000000e+00 -3.485837e-16  9.117530e+00
## CRatio:CpH   0.000000e+00 -3.054826e-14  0.000000e+00  0.000000e+00
##                  I(CpH^2)    CRatio:CpH
## (Intercept) -6.504059e-01  0.000000e+00
## CRatio       0.000000e+00 -3.054826e-14
## CpH         -3.485837e-16  0.000000e+00
## I(CRatio^2)  9.117530e+00  0.000000e+00
## I(CpH^2)     1.093535e+00  0.000000e+00
## CRatio:CpH   0.000000e+00  1.933564e+02
```

```
vcov(model)
```

```
##               (Intercept)        CRatio           CpH   I(CRatio^2)
## (Intercept)  1.221590e+00 -2.247282e-16  1.050224e-16 -6.504059e+01
## CRatio      -2.247282e-16  9.243070e+01 -9.290660e-16  2.938780e-14
## CpH          1.050224e-16 -9.290660e-16  9.243070e-01 -5.382639e-15
## I(CRatio^2) -6.504059e+01  2.938780e-14 -5.382639e-15  1.093535e+04
## I(CpH^2)    -6.504059e-01  1.154629e-16 -3.046570e-16  9.117530e+00
## CRatio:CpH  -2.535431e-15 -4.048999e-14  3.239199e-15  1.975820e-13
##                  I(CpH^2)    CRatio:CpH
## (Intercept) -6.504059e-01 -2.535431e-15
## CRatio       1.154629e-16 -4.048999e-14
## CpH         -3.046570e-16  3.239199e-15
## I(CRatio^2)  9.117530e+00  1.975820e-13
## I(CpH^2)     1.093535e+00  2.642452e-15
## CRatio:CpH   2.642452e-15  1.933564e+02
```

The two matrices appear to be different. Possibly an issue with the precision…

Correlations and variance inflation factors

```
cor(X)
```

```
##             (Intercept)       CRatio          CpH   I(CRatio^2)      I(CpH^2)
## (Intercept)           1           NA           NA            NA            NA
## CRatio               NA 1.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00
## CpH                  NA 0.000000e+00 1.000000e+00  4.578146e-17  3.219493e-16
## I(CRatio^2)          NA 0.000000e+00 4.578146e-17  1.000000e+00 -8.337667e-02
## I(CpH^2)             NA 0.000000e+00 3.219493e-16 -8.337667e-02  1.000000e+00
## CRatio:CpH           NA 1.974107e-16 0.000000e+00  0.000000e+00  0.000000e+00
##               CRatio:CpH
## (Intercept)           NA
## CRatio      1.974107e-16
## CpH         0.000000e+00
## I(CRatio^2) 0.000000e+00
## I(CpH^2)    0.000000e+00
## CRatio:CpH  1.000000e+00
```

```
vif(model)
```

```
##      CRatio         CpH I(CRatio^2)    I(CpH^2)  CRatio:CpH
##       1.000       1.000       1.007       1.007       1.000
```

The variance inflation factors are all very close to 1. This means that any collinearity between predictors is not inflating the variances of the least squares estimates by much.

Confidence intervals for parameter estimates

```
confint(model, level=0.95)
```

```
##                   2.5 %     97.5 %
## (Intercept)   74.049167  78.790243
## CRatio        40.527372  81.767710
## CpH            9.279513  13.403547
## I(CRatio^2) -141.926724 306.643483
## I(CpH^2)     -11.308156  -6.822454
## CRatio:CpH   -47.879384  11.768272
```

Partial F-tests

Fit the partial models

```
m1 <- lm(Yield ~ CRatio + CpH, data=chemical2)
m2 <- lm(Yield ~ CRatio + CpH + (CRatio*CpH), data=chemical2)
m3 <- lm(Yield ~ CRatio + CpH + I(CpH^2), data=chemical2)
```

Do the testing

```
anova(m1, model)
```

```
## Analysis of Variance Table
##
## Model 1: Yield ~ CRatio + CpH
## Model 2: Yield ~ CRatio + CpH + (CRatio * CpH) + I(CRatio^2) + I(CpH^2)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     17 945.20
## 2     14 142.08  3    803.12 26.378 5.078e-06 ***
## ---
```

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

The associated hypotheses are:

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_A : \text{At least one of } \beta_i \text{ non-zero, } i = 3, 4, 5$$

Here we are testing whether the interaction term and the two quadratic terms can be simultaneously removed from the model. The p-value, `5.078e-06`, is less than 0.05. We reject the null hypothesis and conclude that at least one of $\beta_3, \beta_4, \beta_5$ are non-zero.

```r
anova(m2, model)
```

```
## Analysis of Variance Table
##
## Model 1: Yield ~ CRatio + CpH + (CRatio * CpH)
## Model 2: Yield ~ CRatio + CpH + (CRatio * CpH) + I(CRatio^2) + I(CpH^2)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     16 928.09
## 2     14 142.08  2    786.01 38.724 1.971e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The associated hypotheses are:

$$H_0 : \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_A : \text{At least one of } \beta_i \text{ non-zero, } i = 4, 5$$

Here we are testing whether the two quadratic terms can be simultaneously removed from the model. The p-value, `1.971e-06`, is less than 0.05. We reject the null hypothesis and conclude that at least one of β_5, β_6 are non-zero.

```r
anova(m3, model)
```

```
## Analysis of Variance Table
##
## Model 1: Yield ~ CRatio + CpH + I(CpH^2)
## Model 2: Yield ~ CRatio + CpH + (CRatio * CpH) + I(CRatio^2) + I(CpH^2)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     16 165.49
## 2     14 142.08  2    23.406 1.1531 0.3439
```

The associated hypotheses are:

$$H_0 : \beta_3 = \beta_4 = 0 \quad \text{vs} \quad H_A : \text{At least one of } \beta_i \text{ non-zero, } i = 3, 4$$

Here we are testing whether the interaction term and the square of the centred ratio can be simultaneously removed from the model. The p-value, 0.3439, is greater than 0.05. We fail to reject the null hypothesis and conclude that there is insufficient evidence to support that at least one of the two coefficients is non-zero. As such we can remove both the interaction term and the square of the centred ratio.

Low birth weight data

```
# R < 4.0.0
birthwt <- read.table("./lowwt.txt", header=TRUE)
```

```
# R >= 4.0.0
birthwt <- read.table("./lowwt.txt", header=TRUE, stringsAsFactors=TRUE)
```

Data prep

```
class(birthwt$Hospital)
```

```
## [1] "factor"
```

```
birthwt$Hospital
```

```
##  [1] A B A C A B A A C A B A C B A B C A B B C C A C B B C A A A B B B C A B B C
## [39] A C
## Levels: A B C
```

The `Hospital` variable in the data set is a factor. By default, factor levels are ordered alphabetically. **The order of factor levels matters when doing regression!** For the regression models that we will fit in this lab, we want Hospital level "C" to be the reference level. Since "C" is not currently the reference level, we can use the `relevel()` function to make it so.

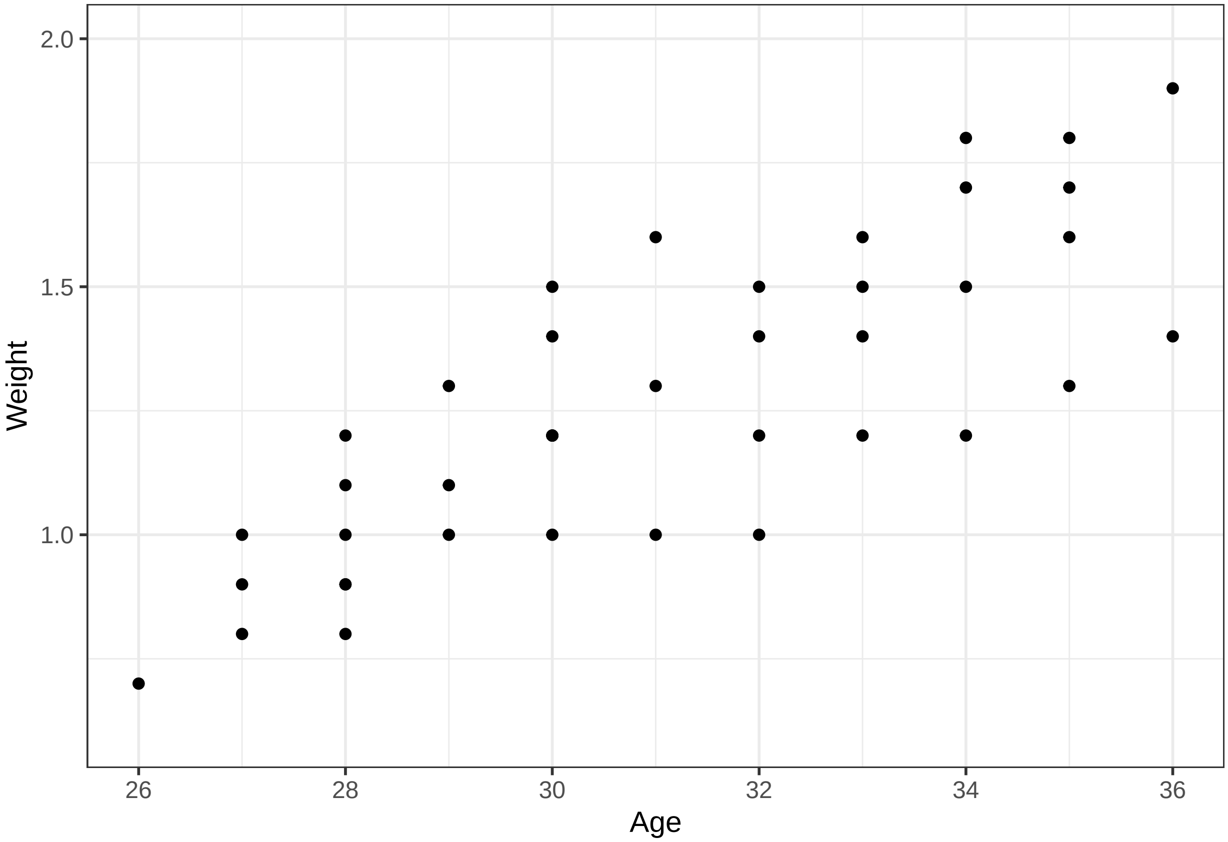
```
(birthwt$Hospital <- relevel(birthwt$Hospital, ref="C"))
```

```
##  [1] A B A C A B A A C A B A C B A B C A B B C C A C B B C A A A B B B C A B B C
## [39] A C
## Levels: C A B
```

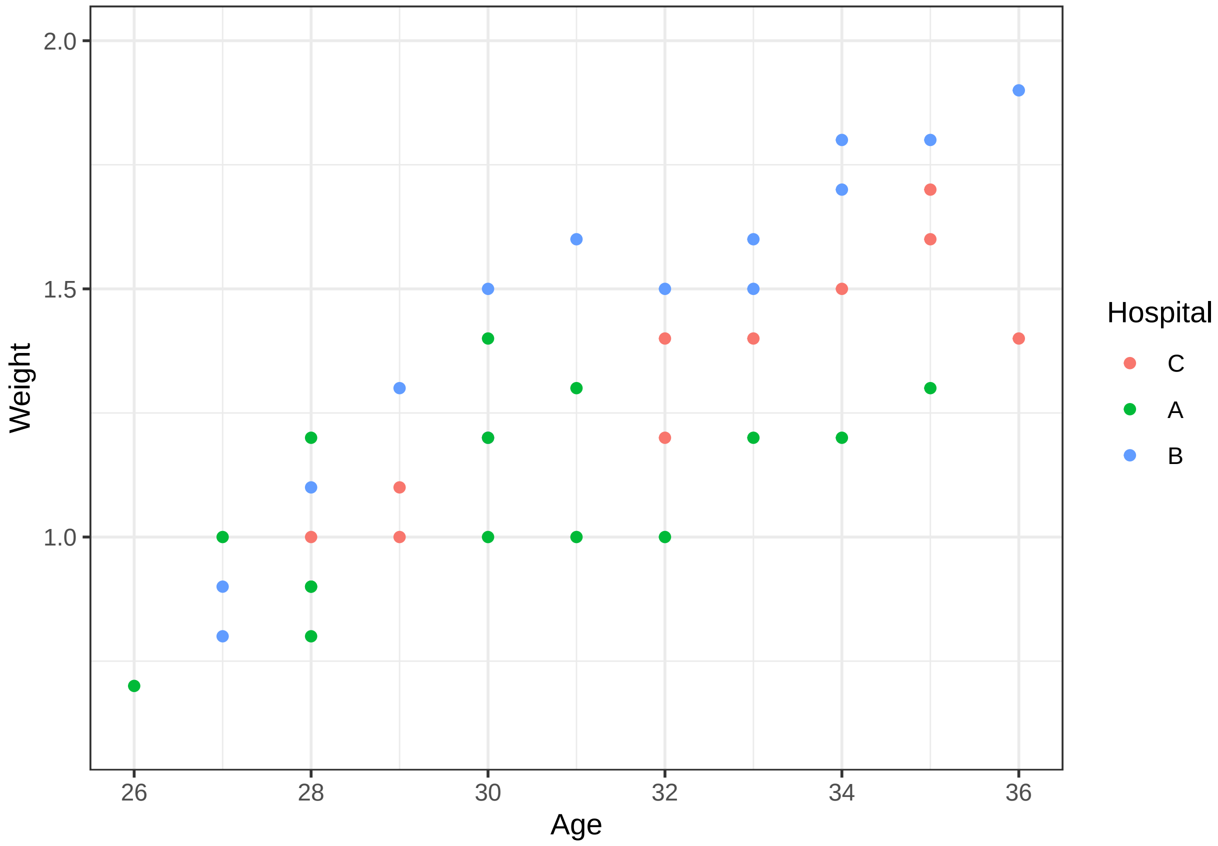
"C" is now the reference level and all other factor levels have been shifted down. We are ready to fit our models.

Scatterplots

```
ggplot(birthwt, aes(x=Age, y=Weight))+
  geom_point()+
  coord_cartesian(ylim=c(0.6, 2))
```



```r
ggplot(birthwt, aes(x=Age, y=Weight, colour=Hospital))+
  geom_point()+
  coord_cartesian(ylim=c(0.6, 2))
```



The legend values are listed in the order of the levels of `Hospital`. They can be re-ordered by adding another layer to the plot using `scale_colour_discrete(breaks=c("A", "B", "C"))`. Don't forget the `+`.

Fit model with `Age`, `Hospital`, and their interaction

```r
model1 <- lm(Weight ~ Age + Hospital + (Age*Hospital), x=TRUE, data=birthwt)
summary(model1)
```

```
##
## Call:
## lm(formula = Weight ~ Age + Hospital + (Age * Hospital), data = birthwt,
##     x = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21714 -0.07487 -0.01513  0.07181  0.32992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.13605    0.49037  -2.317   0.0267 *
## Age           0.07648    0.01523   5.022 1.61e-05 ***
## HospitalA     0.74327    0.63875   1.164   0.2527
## HospitalB    -0.82388    0.62980  -1.308   0.1996
## Age:HospitalA -0.02772    0.02039  -1.359   0.1831
## Age:HospitalB  0.03204    0.01974   1.623   0.1137
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1336 on 34 degrees of freedom
## Multiple R-squared:  0.8336, Adjusted R-squared:  0.8091
## F-statistic: 34.06 on 5 and 34 DF,  p-value: 2.616e-12
```

```
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: Weight
##              Df  Sum Sq Mean Sq  F value    Pr(>F)
## Age           1 2.33141 2.33141 130.7056 3.398e-13 ***
## Hospital      2 0.51854 0.25927  14.5354 2.742e-05 ***
## Age:Hospital  2 0.18760 0.09380   5.2586   0.01024 *
## Residuals    34 0.60646 0.01784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Variance-covariance matrix

```
res1 <- resid(model1)
MSE1 <- sum(res1^2)/model1$df.residual
X1 <- model1$x
XTXinv1 <- solve(t(X1) %*% X1)
(Vhatb1 <- MSE1*XTXinv1)
```

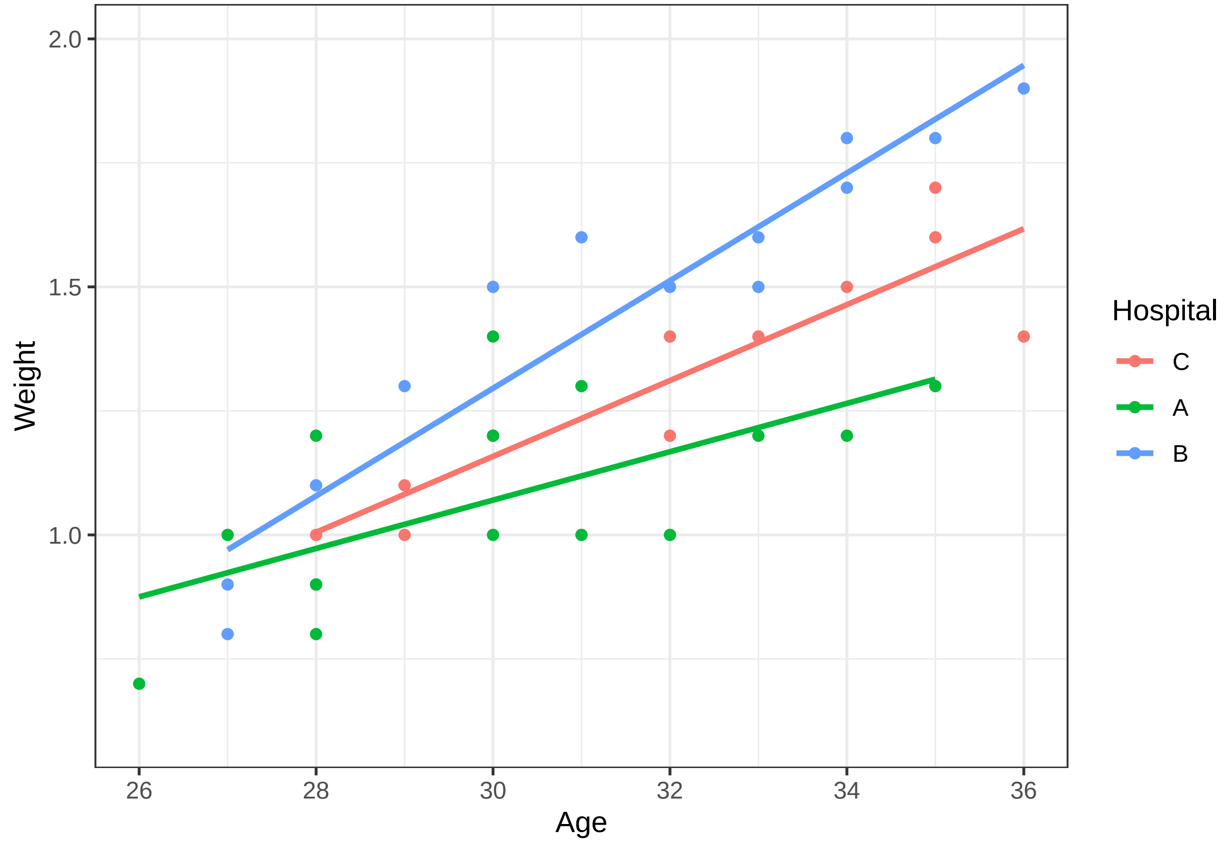
```
##                 (Intercept)           Age    HospitalA    HospitalB
## (Intercept)     0.240463205 -0.0074426577 -0.240463205 -0.240463205
## Age            -0.007442658  0.0002319242  0.007442658  0.007442658
## HospitalA      -0.240463205  0.0074426577  0.408001699  0.240463205
## HospitalB      -0.240463205  0.0074426577  0.240463205  0.396652985
## Age:HospitalA   0.007442658 -0.0002319242 -0.012975341 -0.007442658
## Age:HospitalB   0.007442658 -0.0002319242 -0.007442658 -0.012383022
##                Age:HospitalA Age:HospitalB
## (Intercept)     0.0074426577  0.0074426577
## Age            -0.0002319242 -0.0002319242
## HospitalA      -0.0129753414 -0.0074426577
## HospitalB      -0.0074426577 -0.0123830218
## Age:HospitalA   0.0004159380  0.0002319242
## Age:HospitalB   0.0002319242  0.0003894757
```

```
vcov(model1)
```

```
##                 (Intercept)           Age    HospitalA    HospitalB
## (Intercept)     0.240463205 -0.0074426577 -0.240463205 -0.240463205
## Age            -0.007442658  0.0002319242  0.007442658  0.007442658
## HospitalA      -0.240463205  0.0074426577  0.408001699  0.240463205
## HospitalB      -0.240463205  0.0074426577  0.240463205  0.396652985
## Age:HospitalA   0.007442658 -0.0002319242 -0.012975341 -0.007442658
## Age:HospitalB   0.007442658 -0.0002319242 -0.007442658 -0.012383022
##                Age:HospitalA Age:HospitalB
## (Intercept)     0.0074426577  0.0074426577
## Age            -0.0002319242 -0.0002319242
## HospitalA      -0.0129753414 -0.0074426577
## HospitalB      -0.0074426577 -0.0123830218
## Age:HospitalA   0.0004159380  0.0002319242
## Age:HospitalB   0.0002319242  0.0003894757
```

Visualization of model

```r
birthwt <- transform(birthwt,
                     fitted1 = fitted(model1))
ggplot(birthwt, aes(x=Age, colour=Hospital))+
  geom_point(aes(y=Weight))+
  geom_line(aes(y=fitted1), size=1)+
  coord_cartesian(ylim=c(0.6, 2))
```



Fit model without interaction term

```r
model2 <- lm(Weight ~ Age + Hospital, x=TRUE, data=birthwt)
summary(model2)
```

```
##
## Call:
## lm(formula = Weight ~ Age + Hospital, data = birthwt, x = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29535 -0.10759  0.00870  0.08741  0.33198
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.241228   0.284844  -4.358 0.000105 ***
## Age          0.079755   0.008766   9.098 7.28e-11 ***
## HospitalA   -0.083405   0.061567  -1.355 0.183954
## HospitalB    0.183197   0.060184   3.044 0.004345 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1485 on 36 degrees of freedom
## Multiple R-squared:  0.7821, Adjusted R-squared:  0.7639
## F-statistic: 43.07 on 3 and 36 DF,  p-value: 5.342e-12
```

```
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: Weight
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Age        1 2.33141 2.33141 105.698 2.948e-12 ***
## Hospital   2 0.51854 0.25927  11.754 0.0001178 ***
## Residuals 36 0.79406 0.02206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Variance-covariance matrix

```
res2 <- resid(model2)
MSE2 <- sum(res2^2)/model2$df.residual
X2 <- model2$x
XTXinv2 <- solve(t(X2) %*% X2)
(Vhatb2 <- MSE2*XTXinv2)
```

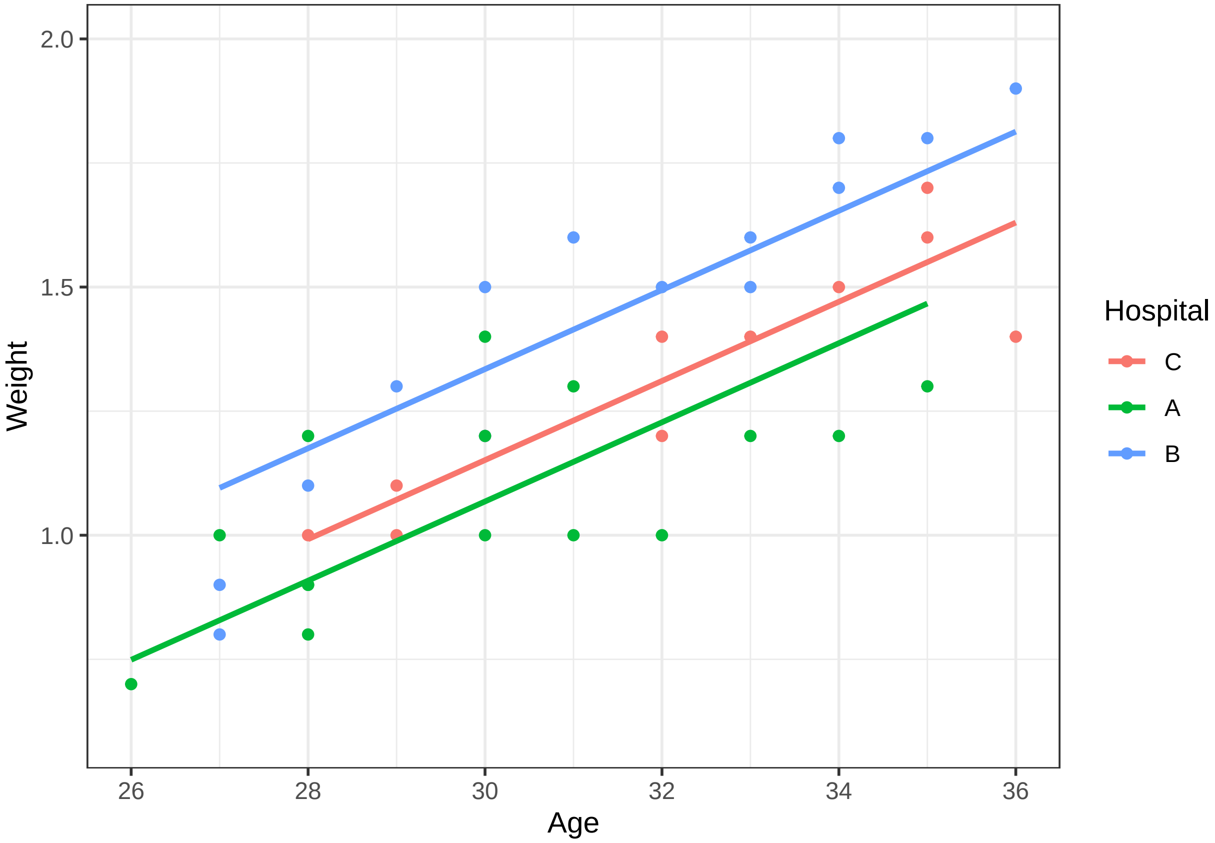
```
##              (Intercept)           Age     HospitalA     HospitalB
## (Intercept)  0.081135939 -2.465831e-03 -0.0069966322 -0.0038145365
## Age         -0.002465831  7.683891e-05  0.0001555406  0.0000563818
## HospitalA   -0.006996632  1.555406e-04  0.0037905202  0.0021193237
## HospitalB   -0.003814536  5.638180e-05  0.0021193237  0.0036220733
```

```
vcov(model2)
```

```
##              (Intercept)           Age     HospitalA     HospitalB
## (Intercept)  0.081135939 -2.465831e-03 -0.0069966322 -0.0038145365
## Age         -0.002465831  7.683891e-05  0.0001555406  0.0000563818
## HospitalA   -0.006996632  1.555406e-04  0.0037905202  0.0021193237
## HospitalB   -0.003814536  5.638180e-05  0.0021193237  0.0036220733
```

Visualization of model

```
birthwt <- transform(birthwt,
                     fitted2 = fitted(model2))
ggplot(birthwt, aes(x=Age, colour=Hospital))+
  geom_point(aes(y=Weight))+
  geom_line(aes(y=fitted2), size=1)+
  coord_cartesian(ylim=c(0.6, 2))
```



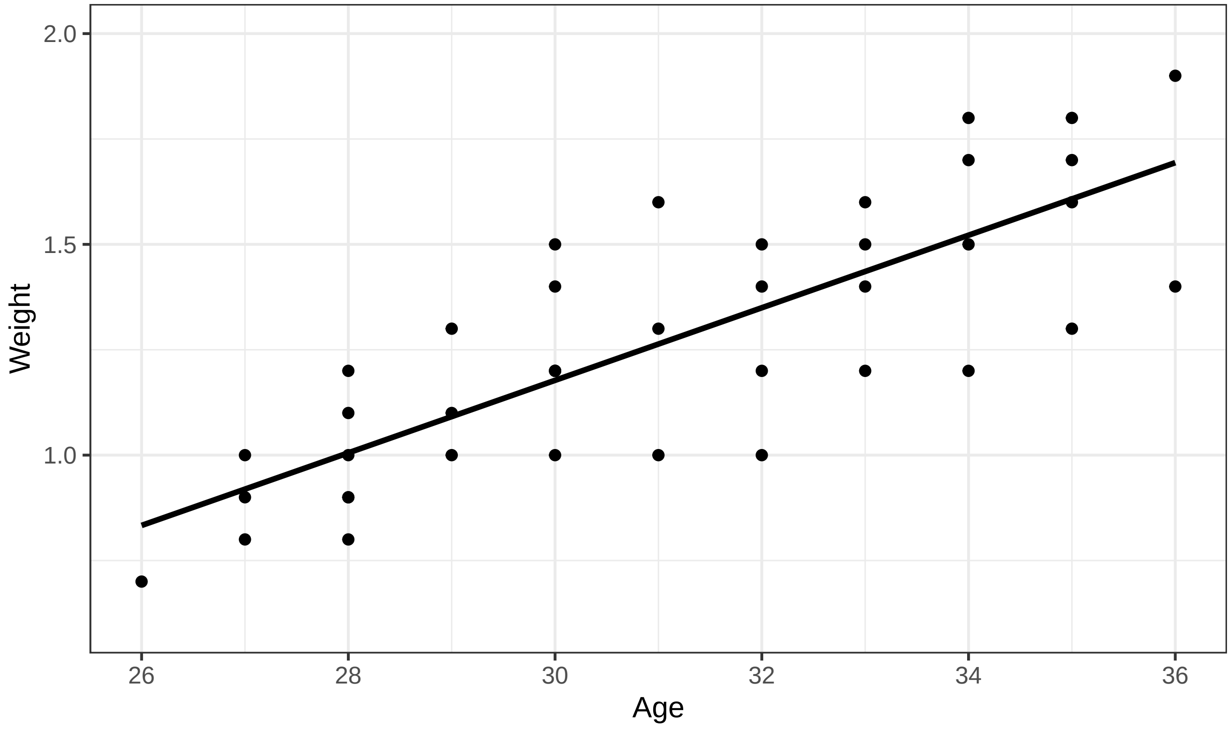
Fit simple linear regression model

```
model3 <- lm(Weight ~ Age, data=birthwt)
summary(model3)
```

```
##
## Call:
## lm(formula = Weight ~ Age, data = birthwt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34961 -0.12278  0.01555  0.15388  0.33645
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.40436    0.32685  -4.297 0.000116 ***
## Age          0.08606    0.01048   8.216 5.95e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1859 on 38 degrees of freedom
## Multiple R-squared:  0.6398, Adjusted R-squared:  0.6303
## F-statistic: 67.49 on 1 and 38 DF,  p-value: 5.953e-10
```

Visualization of model

```
birthwt <- transform(birthwt,
                     fitted3 = fitted(model3))
ggplot(birthwt, aes(x=Age))+
  geom_point(aes(y=Weight))+
  geom_line(aes(y=fitted3), size=1)+
  coord_cartesian(ylim=c(0.6, 2))
```



Model comparisons

```r
anova(model2, model1)
```

```
## Analysis of Variance Table
##
## Model 1: Weight ~ Age + Hospital
## Model 2: Weight ~ Age + Hospital + (Age * Hospital)
##   Res.Df     RSS Df Sum of Sq      F  Pr(>F)
## 1     36 0.79406
## 2     34 0.60646  2    0.1876 5.2586 0.01024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The associated hypotheses are:

$$H_0 : \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_A : \text{At least one of } \beta_i \text{ non-zero, } i = 4, 5$$

We are testing whether we can remove the interaction terms. Since the p-value, 0.01204, is less than 0.05, we reject the null hypothesis. As such, we should not remove the interaction terms from the full model.

```r
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: Weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Age        1 2.3314 2.33141  67.495 5.953e-10 ***
## Residuals 38 1.3126 0.03454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The associated hypotheses are:

$$H_0 : \beta_1 = 0, \quad \text{vs} \quad H_A : \beta_1 \neq 0$$

This is looking at model usefulness in the simple linear regression model (assumes that the assumptions of linear regression are satisfied). Since the p-value, `5.963e-10`, is less than 0.05, we reject the null hypothesis. We conclude that our model is useful.

```r
anova(model3, model1)
```

```
## Analysis of Variance Table
##
## Model 1: Weight ~ Age
## Model 2: Weight ~ Age + Hospital + (Age * Hospital)
##   Res.Df     RSS Df Sum of Sq     F    Pr(>F)
## 1     38 1.31259
## 2     34 0.60646  4   0.70613 9.897 2.022e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The associated hypotheses are:

$$H_0 : \text{All } \beta_i = 0 \quad \text{vs} \quad H_A : \text{At least one of } \beta_i \text{ non-zero, } i = 2, 3, 4, 5$$

We are testing whether we can remove the `Hospital` terms and the interaction of `Age` with `Hospital`. Since the p-value, `2.022e-05`, is less than 0.05, we reject the null hypothesis. As such, we cannot simultaneously remove the effect of `Hospital` and the interaction of `Age` and `Hospital` from the full model.

```
anova(model3, model2)
```

```
## Analysis of Variance Table
##
## Model 1: Weight ~ Age
## Model 2: Weight ~ Age + Hospital
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     38 1.31259
## 2     36 0.79406  2   0.51854 11.754 0.0001178 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The associated hypotheses are:

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_A : \text{At least one of } \beta_i \text{ non-zero}, \, i = 2, 3$$

We are testing whether we can remove the `Hospital` terms from the model. Since the p-value, 0.0001178, is less than 0.05, we reject the null hypothesis. We conclude that we cannot remove the `Hospital` terms from the second model.

Predictions

It was suggested in the lab instructions to create all the variables prior to fitting the model, such as the indicator variables and the interaction terms. While this is fine to do, if you need to do any sort of prediction with your model, you will need to supply values for **ALL** of these variables. This is because R doesn't know that these variables originated from one another.

If we create the variables within the formula specification, R knows that these variables are related. As such, when we do predictions we can supply fewer values. Especially for models with many terms consisting of interactions and indicator variables, entering all the values manually to obtain a prediction leaves more room for human error.

```
predict(model1, newdata=data.frame(Age=27, Hospital="A"))
```

```
##         1
## 0.9237964
```

However, for models with centred variables such as the one we made for the `chemical` data, I would recommend doing the centring prior to fitting the model. Otherwise predictions become tricky to obtain. In addition, it is important to remember that when predicting with centred variables, that you supply a centred value rather than a raw value!