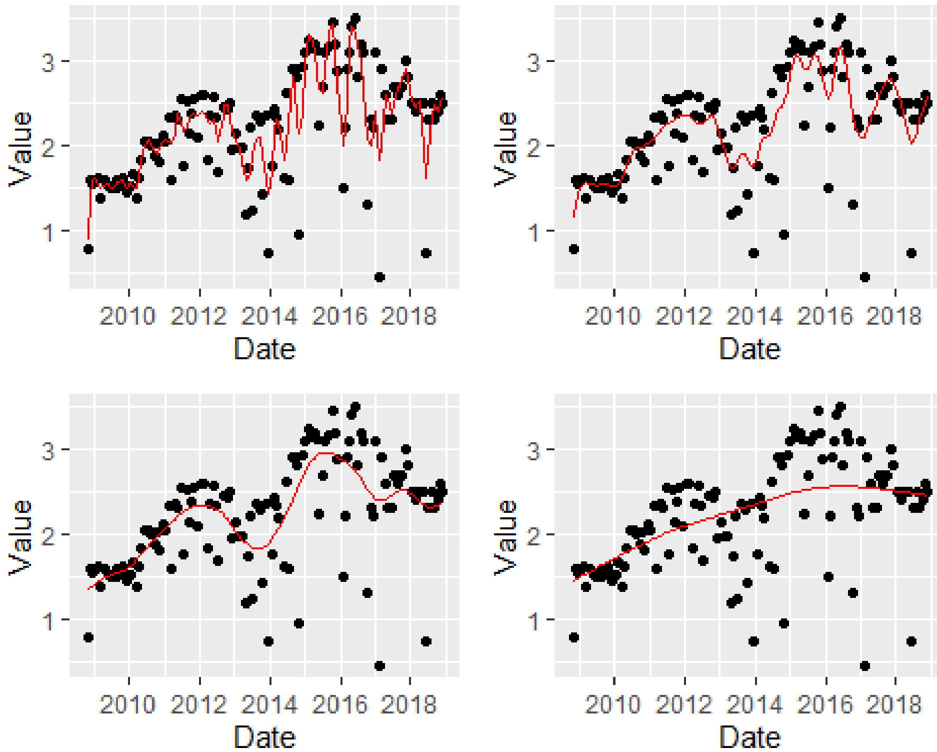
Assignment6

2023-05-29

```
library(gam)

## Loading required package: splines

## Loading required package: foreach

## Loaded gam 1.22-2

library(ggplot2)
library(gridExtra)
ccc <- read.csv("CCC05.csv")
ecan <- read.csv("ECAN93.csv")
ccc$Date <- as.Date(ccc$Date, format="%d/%m/%Y")
mod1 <- gam(Value ~ s(Date, spar = 0.2), data=ccc)
mod2 <- gam(Value ~ s(Date, spar = 0.4), data=ccc)
mod3 <- gam(Value ~ s(Date, spar = 0.63), data=ccc)
mod4 <- gam(Value ~ s(Date, spar = 1), data=ccc)
ccc$predict1 <- predict(mod1)
ccc$predict2 <- predict(mod2)
ccc$predict3 <- predict(mod3)
ccc$predict4 <- predict(mod4)

plot1 <- ggplot(ccc, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict1),                     col='red')
plot2 <- ggplot(ccc, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict2),                     col='red')
plot3 <- ggplot(ccc, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict3),                    col='red')
plot4 <- ggplot(ccc, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict4),                     col='red')

grid.arrange(plot1, plot2, plot3, plot4, nrow=2)
```




```
summary(mod3)

##
## Call: gam(formula = Value ~ s(Date, spar = 0.63), data = ccc)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.94583 -0.12562  0.09761  0.25403  0.74684
##
## (Dispersion Parameter for gaussian family taken to be 0.2292)
##
##     Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 25.3442 on 110.5726 degrees of freedom
## AIC: 181.6184
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##                            Df Sum Sq Mean Sq F value    Pr(>F)
## s(Date, spar = 0.63)   1.00 12.219 12.2190  53.309 4.654e-11 ***
## Residuals            110.57 25.344  0.2292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                        Npar Df Npar F     Pr(F)
## (Intercept)
## s(Date, spar = 0.63)      10.4 5.0305 3.79e-06 ***
```

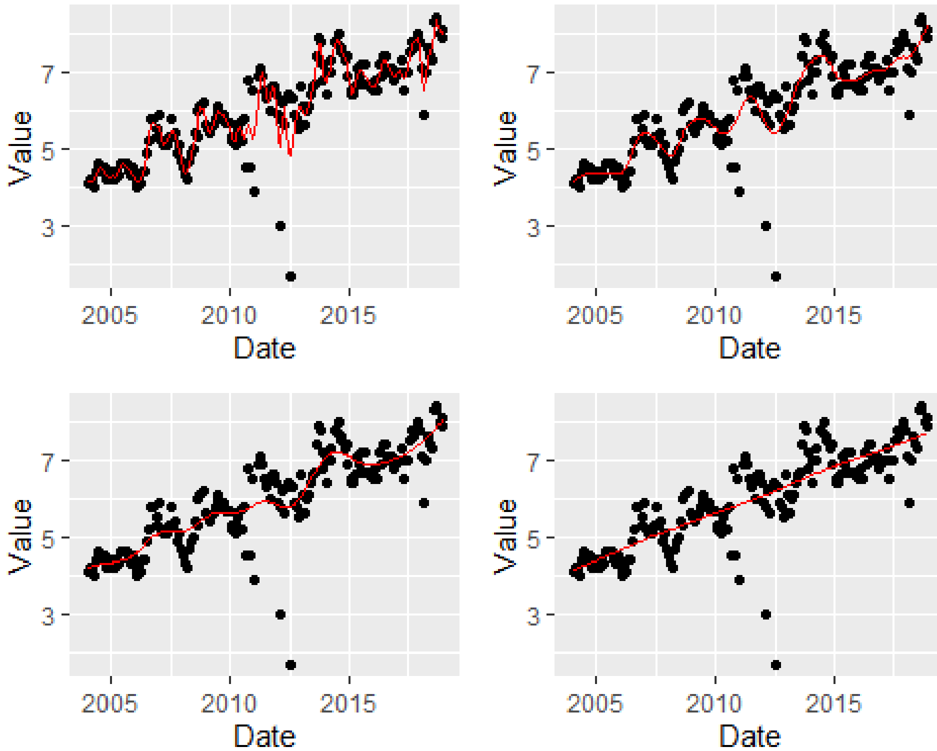
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I will not be using AIC or p-values to determine the best models. AIC values are not relevant because an overfitted model with have low AIC. P-values are not relevant because an insignificant p-value means a model may be over smoothed while a significant p value does not tell us if the model is over fitted or a good fit.

Mod1 and Mod2 over fit the data as the model captures features that may not actually be present, this causes a lot of noise in the data. Mod4 over smooths the data as the model does not capture the important features. Mod3 has a spar of 0.63. This looks a good fit as it only captures the significant features in the model without capturing trends they may not actually be present in the data.

Mod3 shows the Nitrate levels in th Christchurch City river steadily increases over time. There appears to be a seasonal trend as there are clear peaks in 2012 and mid 2015. There are clear troughs in mid 2013 and mid 2018.

```
ecan$Date <- as.Date(ecan$Date, format="%d/%m/%Y")
mod5 <- gam(Value ~ s(Date, spar = 0.2), data=ecan)
mod6 <- gam(Value ~ s(Date, spar = 0.55), data=ecan)
mod7 <- gam(Value ~ s(Date, spar = 0.7), data=ecan)
mod8 <- gam(Value ~ s(Date, spar = 1), data=ecan)

ecan$predict5 <- predict(mod5)
ecan$predict6 <- predict(mod6)
ecan$predict7 <- predict(mod7)
ecan$predict8 <- predict(mod8)

plot5 <- ggplot(ecan, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict5),                    col='red')
plot6 <- ggplot(ecan, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict6),                    col='red')
plot7 <- ggplot(ecan, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict7),                    col='red')
plot8 <- ggplot(ecan, aes(x=Date, y=Value)) + geom_point() +
geom_line(aes(x=Date, y=predict8),                    col='red')
grid.arrange(plot5, plot6, plot7, plot8, nrow=2)
```



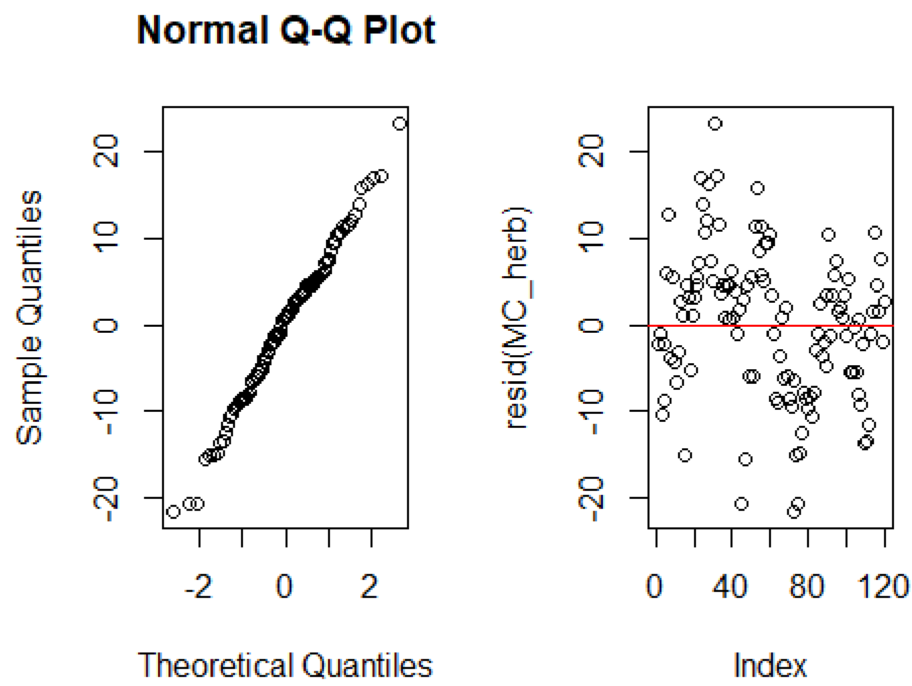

```
summary(mod2)

##
## Call: gam(formula = Value ~ s(Date, spar = 0.4), data = ccc)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.65041 -0.09580  0.07344  0.23888  0.99659
##
## (Dispersion Parameter for gaussian family taken to be 0.2128)
##
##     Null Deviance: 49.5862 on 122 degrees of freedom
## Residual Deviance: 19.8108 on 93.1011 degrees of freedom
## AIC: 186.2642
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##                          Df Sum Sq Mean Sq F value    Pr(>F)
## s(Date, spar = 0.4)  1.000 12.219 12.2190  57.423 2.544e-11 ***
## Residuals           93.101 19.811  0.2128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                     Npar Df Npar F     Pr(F)
## (Intercept)
## s(Date, spar = 0.4)    27.9 2.9573 5.109e-05 ***
```

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

Mod5 overfits the data, it captures the main feautes in the data but it also captures features that may not actually exist. Mod7 and Mod8 both over smooth the data as the main features are not captured. Mod6 has a spar of 0.55. This appears to be the best model as it captures the main features of the data without creating noise by capturing features that may not exist.

Mod6 shows the Nitrate levels in the Canterbury Region river are increasing over time on average. There appears to be a seasonal effect as the Nirtate levels in the river have a clear peak followed by a trough. This may be due to changing conditions in the environment such as temperature.

```
library(readxl)
library(agricolae)
library(multcomp)

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

herb <- read_excel("Herbicides.xlsx")

MC_herb <- aov(Grass_percent ~ Herbicide, herb)
summary(MC_herb)

##              Df Sum Sq Mean Sq F value   Pr(>F)
## Herbicide     9   3092   343.5   4.412 6.09e-05 ***
## Residuals   110   8564    77.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

par(mfrow=(c(1,2)))
qqnorm(MC_herb$residuals)
plot(resid(MC_herb))
abline(0,0, col='red')
```



The residuals are fitted to a relatively straight line on the normal Q-Q plot and the residual plot shows constant variance. This means we can assume our data is normally distributed.

```
mse <-sum(MC_herb$residuals*MC_herb$residuals)/MC_herb$df.residual
LSD.test(herb$Grass_percent, herb$Herbicide, MC_herb$df.residual, mse,
console = TRUE)

##
## Study: herb$Grass_percent ~ herb$Herbicide
##
## LSD t Test for herb$Grass_percent
##
## Mean Square Error:  77.8534
##
## herb$Herbicide,  means and individual ( 95 %) CI
##
##                          herb.Grass_percent       std  r      LCL      UCL
Min
## Aminopyralid                       63.44375  9.913055 12 58.39597 68.49153
49.875
## Aminopyralid+triclopyr             62.89583  8.645617 12 57.84805 67.94361
52.500
## Chlorsulfuron                      52.77083  5.158244 12 47.72305 57.81861
44.500
## Flumetsulam                        58.09375  6.201202 12 53.04597 63.14153
49.500
## MCPA                               58.31875  8.093657 12 53.27097 63.36653
45.750
## MCPB                               55.11458 10.260590 12 50.06680 60.16236
```

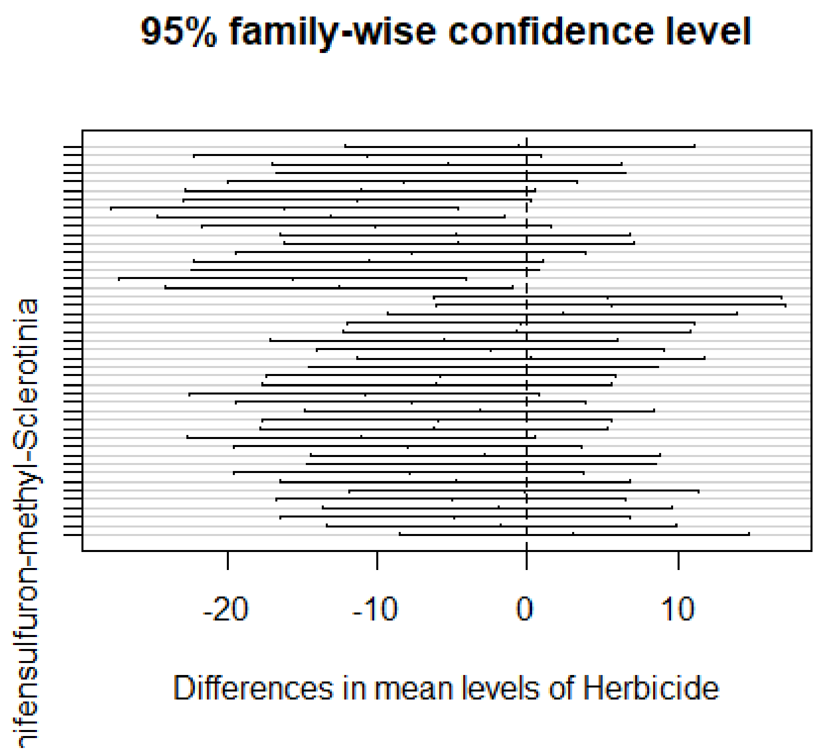
```
33.500
## MCPB+bentazone                   52.29167  8.893201 12 47.24389 57.33945
36.750
## Nil                              52.04167  7.303551 12 46.99389 57.08945
40.375
## Sclerotinia                      47.19792 12.355696 12 42.15014 52.24570
26.500
## Thifensulfuron-methyl            50.30208  9.196476 12 45.25430 55.34986
29.500
##                           Max
## Aminopyralid          86.75
## Aminopyralid+triclopyr 80.00
## Chlorsulfuron         60.25
## Flumetsulam           70.75
## MCPA                  69.75
## MCPB                  72.00
## MCPB+bentazone        64.25
## Nil                   63.50
## Sclerotinia           63.50
## Thifensulfuron-methyl  64.25
##
## Alpha: 0.05 ; DF Error: 110
## Critical Value of t: 1.981765
##
## least Significant Difference: 7.138638
##
## Treatments with the same letter are not significantly different.
##
##                          herb$Grass_percent groups
## Aminopyralid                       63.44375      a
## Aminopyralid+triclopyr             62.89583      a
## MCPA                               58.31875     ab
## Flumetsulam                        58.09375     ab
## MCPB                               55.11458     bc
## Chlorsulfuron                      52.77083    bcd
## MCPB+bentazone                     52.29167    bcd
## Nil                                52.04167    bcd
## Thifensulfuron-methyl              50.30208     cd
## Sclerotinia                        47.19792      d
```

The pairs with significant differences according to the LSD test are the pairs which do not have any of the letters. This means that the pairs of herbicedes have a difference in grass_percent of atleast 7.14. In this data there are 17 pairs of herbicides that have significant differences in grass_percent according to the LSD test.

```
pairwise.t.test(herb$Grass_percent, herb$Herbicide, p.adj =
"bonferroni",console=TRUE)
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  herb$Grass_percent and herb$Herbicide
##
##                         Aminopyralid Aminopyralid+triclopyr Chlorsulfuron
## Aminopyralid+triclopyr 1.00000      -                      -
## Chlorsulfuron          0.16810      0.26332                -
## Flumetsulam            1.00000      1.00000                1.00000
## MCPA                   1.00000      1.00000                1.00000
## MCPB                   1.00000      1.00000                1.00000
## MCPB+bentazone         0.11201      0.17800                1.00000
## Nil                    0.09018      0.14438                1.00000
## Sclerotinia            0.00073      0.00133                1.00000
## Thifensulfuron-methyl  0.01824      0.03068                1.00000
##                         Flumetsulam MCPA    MCPB    MCPB+bentazone Nil
## Aminopyralid+triclopyr -           -       -       -              -
## Chlorsulfuron          -           -       -       -              -
## Flumetsulam            -           -       -       -              -
## MCPA                   1.00000     -       -       -              -
## MCPB                   1.00000     1.00000 -       -              -
## MCPB+bentazone         1.00000     1.00000 1.00000 -              -
## Nil                    1.00000     1.00000 1.00000 1.00000        -
## Sclerotinia            0.13938     0.11505 1.00000 1.00000        1.00000
## Thifensulfuron-methyl  1.00000     1.00000 1.00000 1.00000        1.00000
##                         Sclerotinia
## Aminopyralid+triclopyr -
## Chlorsulfuron          -
## Flumetsulam            -
## MCPA                   -
## MCPB                   -
## MCPB+bentazone         -
## Nil                    -
## Sclerotinia            -
## Thifensulfuron-methyl  1.00000
##
## P value adjustment method: bonferroni
```

```
herbHSD <- TukeyHSD(aov(Grass_percent ~ Herbicide, herb))
   plot(herbHSD)
```




```
print(herbHSD)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Grass_percent ~ Herbicide, data = herb)
##
## $Herbicide
##                                                    diff        lwr
upr
## Aminopyralid+triclopyr-Aminopyralid           -0.5479167 -12.181524
11.0856909
## Chlorsulfuron-Aminopyralid                   -10.6729167 -22.306524
0.9606909
## Flumetsulam-Aminopyralid                      -5.3500000 -16.983608
6.2836076
## MCPA-Aminopyralid                             -5.1250000 -16.758608
6.5086076
## MCPB-Aminopyralid                             -8.3291667 -19.962774
3.3044409
## MCPB+bentazone-Aminopyralid                  -11.1520833 -22.785691
0.4815243
## Nil-Aminopyralid                             -11.4020833 -23.035691
0.2315243
## Sclerotinia-Aminopyralid                     -16.2458333 -27.879441 -
4.6122257
## Thifensulfuron-methyl-Aminopyralid           -13.1416667 -24.775274 -
```

```
1.5080591
## Chlorsulfuron-Aminopyralid+triclopyr         -10.1250000 -21.758608
1.5086076
## Flumetsulam-Aminopyralid+triclopyr            -4.8020833 -16.435691
6.8315243
## MCPA-Aminopyralid+triclopyr                   -4.5770833 -16.210691
7.0565243
## MCPB-Aminopyralid+triclopyr                   -7.7812500 -19.414858
3.8523576
## MCPB+bentazone-Aminopyralid+triclopyr        -10.6041667 -22.237774
1.0294409
## Nil-Aminopyralid+triclopyr                   -10.8541667 -22.487774
0.7794409
## Sclerotinia-Aminopyralid+triclopyr           -15.6979167 -27.331524 -
4.0643091
## Thifensulfuron-methyl-Aminopyralid+triclopyr -12.5937500 -24.227358 -
0.9601424
## Flumetsulam-Chlorsulfuron                      5.3229167  -6.310691
16.9565243
## MCPA-Chlorsulfuron                             5.5479167  -6.085691
17.1815243
## MCPB-Chlorsulfuron                             2.3437500  -9.289858
13.9773576
## MCPB+bentazone-Chlorsulfuron                  -0.4791667 -12.112774
11.1544409
## Nil-Chlorsulfuron                             -0.7291667 -12.362774
10.9044409
## Sclerotinia-Chlorsulfuron                     -5.5729167 -17.206524
6.0606909
## Thifensulfuron-methyl-Chlorsulfuron           -2.4687500 -14.102358
9.1648576
## MCPA-Flumetsulam                               0.2250000 -11.408608
11.8586076
## MCPB-Flumetsulam                              -2.9791667 -14.612774
8.6544409
## MCPB+bentazone-Flumetsulam                    -5.8020833 -17.435691
5.8315243
## Nil-Flumetsulam                               -6.0520833 -17.685691
5.5815243
## Sclerotinia-Flumetsulam                      -10.8958333 -22.529441
0.7377743
## Thifensulfuron-methyl-Flumetsulam             -7.7916667 -19.425274
3.8419409
## MCPB-MCPA                                     -3.2041667 -14.837774
8.4294409
## MCPB+bentazone-MCPA                           -6.0270833 -17.660691
5.6065243
## Nil-MCPA                                      -6.2770833 -17.910691
5.3565243
## Sclerotinia-MCPA                             -11.1208333 -22.754441
```

```
0.5127743
## Thifensulfuron-methyl-MCPA                        -8.0166667 -19.650274
3.6169409
## MCPB+bentazone-MCPB                               -2.8229167 -14.456524
8.8106909
## Nil-MCPB                                          -3.0729167 -14.706524
8.5606909
## Sclerotinia-MCPB                                  -7.9166667 -19.550274
3.7169409
## Thifensulfuron-methyl-MCPB                        -4.8125000 -16.446108
6.8211076
## Nil-MCPB+bentazone                                -0.2500000 -11.883608
11.3836076
## Sclerotinia-MCPB+bentazone                        -5.0937500 -16.727358
6.5398576
## Thifensulfuron-methyl-MCPB+bentazone              -1.9895833 -13.623191
9.6440243
## Sclerotinia-Nil                                   -4.8437500 -16.477358
6.7898576
## Thifensulfuron-methyl-Nil                         -1.7395833 -13.373191
9.8940243
## Thifensulfuron-methyl-Sclerotinia                  3.1041667  -8.529441
14.7377743
##                                                      p adj
## Aminopyralid+triclopyr-Aminopyralid              1.0000000
## Chlorsulfuron-Aminopyralid                       0.1012182
## Flumetsulam-Aminopyralid                         0.8952159
## MCPA-Aminopyralid                                0.9175099
## MCPB-Aminopyralid                                0.3896947
## MCPB+bentazone-Aminopyralid                      0.0719191
## Nil-Aminopyralid                                 0.0596955
## Sclerotinia-Aminopyralid                         0.0006647
## Thifensulfuron-methyl-Aminopyralid               0.0142167
## Chlorsulfuron-Aminopyralid+triclopyr             0.1457962
## Flumetsulam-Aminopyralid+triclopyr               0.9436704
## MCPA-Aminopyralid+triclopyr                      0.9580631
## MCPB-Aminopyralid+triclopyr                      0.4901410
## MCPB+bentazone-Aminopyralid+triclopyr            0.1061242
## Nil-Aminopyralid+triclopyr                       0.0891595
## Sclerotinia-Aminopyralid+triclopyr               0.0011879
## Thifensulfuron-methyl-Aminopyralid+triclopyr 0.0228998
## Flumetsulam-Chlorsulfuron                        0.8980798
## MCPA-Chlorsulfuron                               0.8727875
## MCPB-Chlorsulfuron                               0.9997063
## MCPB+bentazone-Chlorsulfuron                     1.0000000
## Nil-Chlorsulfuron                                1.0000000
## Sclerotinia-Chlorsulfuron                        0.8697678
## Thifensulfuron-methyl-Chlorsulfuron              0.9995511
## MCPA-Flumetsulam                                 1.0000000
## MCPB-Flumetsulam                                 0.9980023
```

```
## MCPB+bentazone-Flumetsulam                        0.8401778
## Nil-Flumetsulam                                   0.8041383
## Sclerotinia-Flumetsulam                           0.0865607
## Thifensulfuron-methyl-Flumetsulam                 0.4881667
## MCPB-MCPA                                         0.9965119
## MCPB+bentazone-MCPA                               0.8079096
## Nil-MCPA                                          0.7686463
## Sclerotinia-MCPA                                  0.0735856
## Thifensulfuron-methyl-MCPA                        0.4460368
## MCPB+bentazone-MCPB                               0.9986881
## Nil-MCPB                                          0.9974634
## Sclerotinia-MCPB                                  0.4646294
## Thifensulfuron-methyl-MCPB                        0.9429303
## Nil-MCPB+bentazone                                1.0000000
## Sclerotinia-MCPB+bentazone                        0.9203383
## Thifensulfuron-methyl-MCPB+bentazone              0.9999248
## Sclerotinia-Nil                                   0.9406693
## Thifensulfuron-methyl-Nil                         0.9999760
## Thifensulfuron-methyl-Sclerotinia                 0.9972591
```

```
  Tukeyaov <- aov(Grass_percent ~ Herbicide, herb)
  summary(Tukeyaov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Herbicide     9   3092   343.5   4.412 6.09e-05 ***
## Residuals   110   8564    77.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

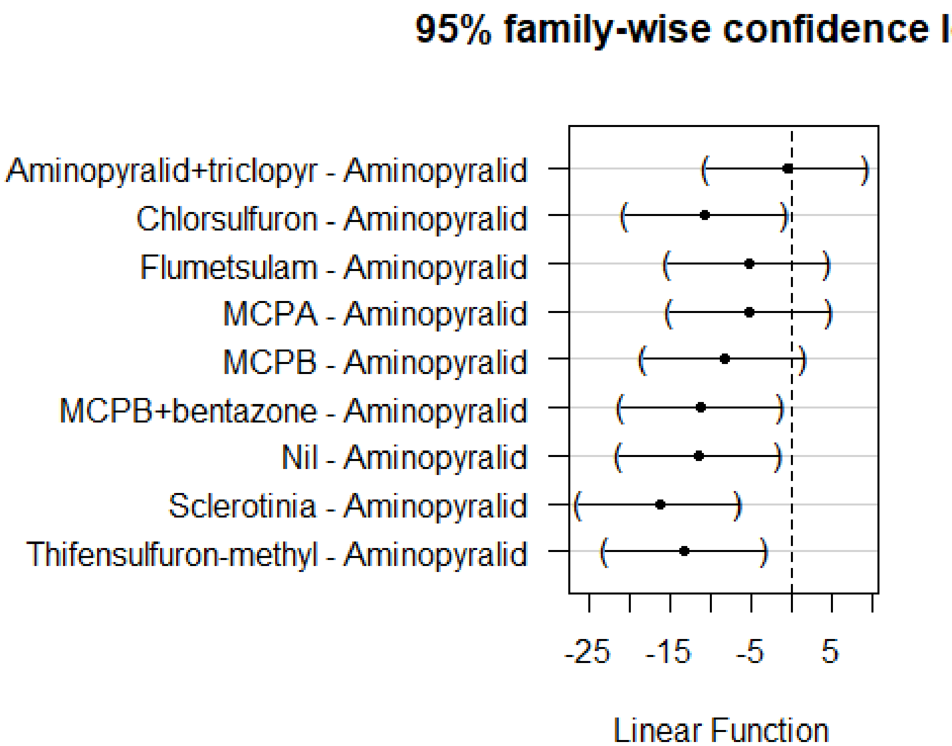
```
   herbHSD <- TukeyHSD(Tukeyaov)
```

```
 herb$Trta <- as.factor(herb$Herbicide)
 Dunnetaov <- aov(Grass_percent ~ Trta, herb)
 test.dunnett=glht(Dunnetaov,linfct=mcp(Trta="Dunnett"))
 confint(test.dunnett)
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Dunnett Contrasts
##
##
## Fit: aov(formula = Grass_percent ~ Trta, data = herb)
##
## Quantile = 2.7315
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                                            Estimate lwr      upr
## Aminopyralid+triclopyr - Aminopyralid == 0  -0.5479 -10.3873   9.2914
```

```
## Chlorsulfuron - Aminopyralid == 0           -10.6729 -20.5123  -0.8336
## Flumetsulam - Aminopyralid == 0              -5.3500 -15.1894   4.4894
## MCPA - Aminopyralid == 0                     -5.1250 -14.9644   4.7144
## MCPB - Aminopyralid == 0                     -8.3292 -18.1685   1.5102
## MCPB+bentazone - Aminopyralid == 0          -11.1521 -20.9914  -1.3127
## Nil - Aminopyralid == 0                     -11.4021 -21.2414  -1.5627
## Sclerotinia - Aminopyralid == 0             -16.2458 -26.0852  -6.4065
## Thifensulfuron-methyl - Aminopyralid == 0  -13.1417 -22.9810  -3.3023
```

```
op <- par()
par(mar=c(4,15,4,2))
plot(test.dunnett)
```




```
par(op)

## Warning in par(op): graphical parameter "cin" cannot be set

## Warning in par(op): graphical parameter "cra" cannot be set

## Warning in par(op): graphical parameter "csi" cannot be set

## Warning in par(op): graphical parameter "cxy" cannot be set

## Warning in par(op): graphical parameter "din" cannot be set

## Warning in par(op): graphical parameter "page" cannot be set
```

The Bonferroni and Tukey adjustments show there are 4 pairs with significant p-values meaning there are 4 pairs of herbicides that have significant differences in grass_percent. These pairs are (Aminopyralid, Sclerotinia), (Aminopyralid, Thifensulfuron-methyl), (Aminopyralid+triclopyr, Sclerotini) and (Aminopyralid+triclopyr, Thifensulfuron-methyl). This means we can be confident there is a significant difference between these herbicides.

The Dunnett method shows that Aminipyralid alone has a significantly different impact on grass_percent when compared to the 5 herbicides Thifensulfuron-methyl, Sclerotinia, Nil, MCPB+bentazone and Chlorsulfuron. This means there could possibly be a significant difference between these herbicides.

The LSD tests claims 17 significant differences while Bonferroni and Tukey only claim 4 and the Dunnett claims 5 for Aminopyralid herbicide. I would use the Bonferroni, Tukey and Dunnett adjustments over the LSD test because the LSD test claims many significant differences between herbicides which the other methods do not.