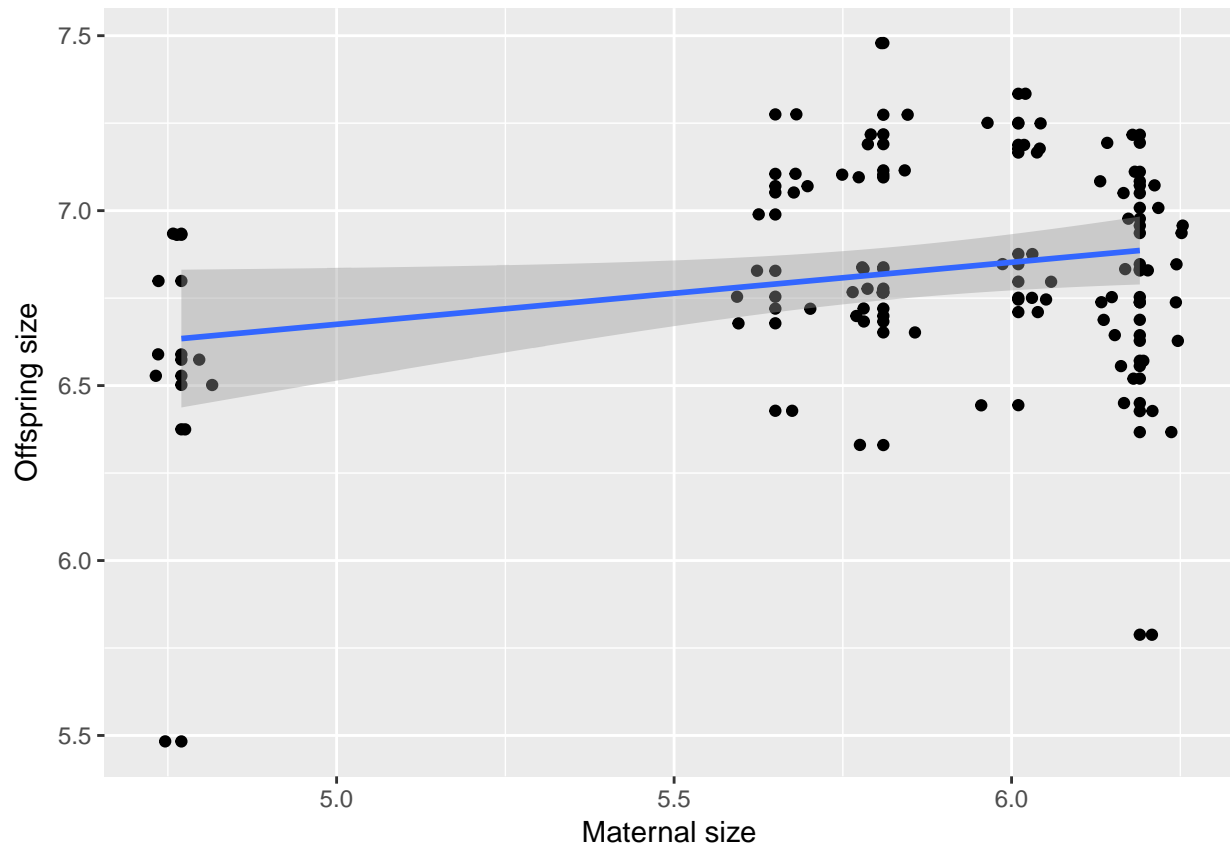


F2 maternal effect

Erik Parker

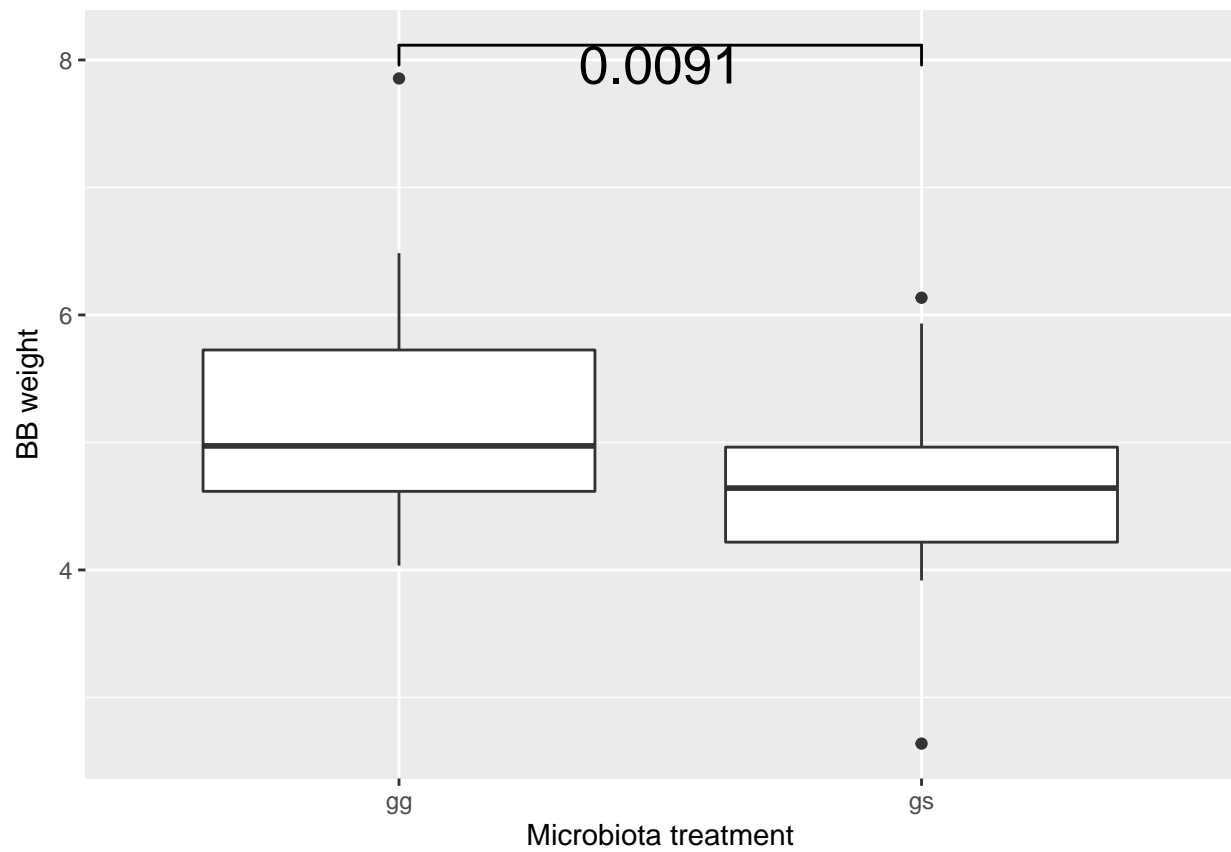
May 11, 2018

```
ggplot(f2.validate, aes(x = Maternal.size, y = Scope.size)) + geom_point() +  
  geom_jitter() + geom_smooth(method = "lm") + labs(y = "Offspring size",  
  x = "Maternal size")
```

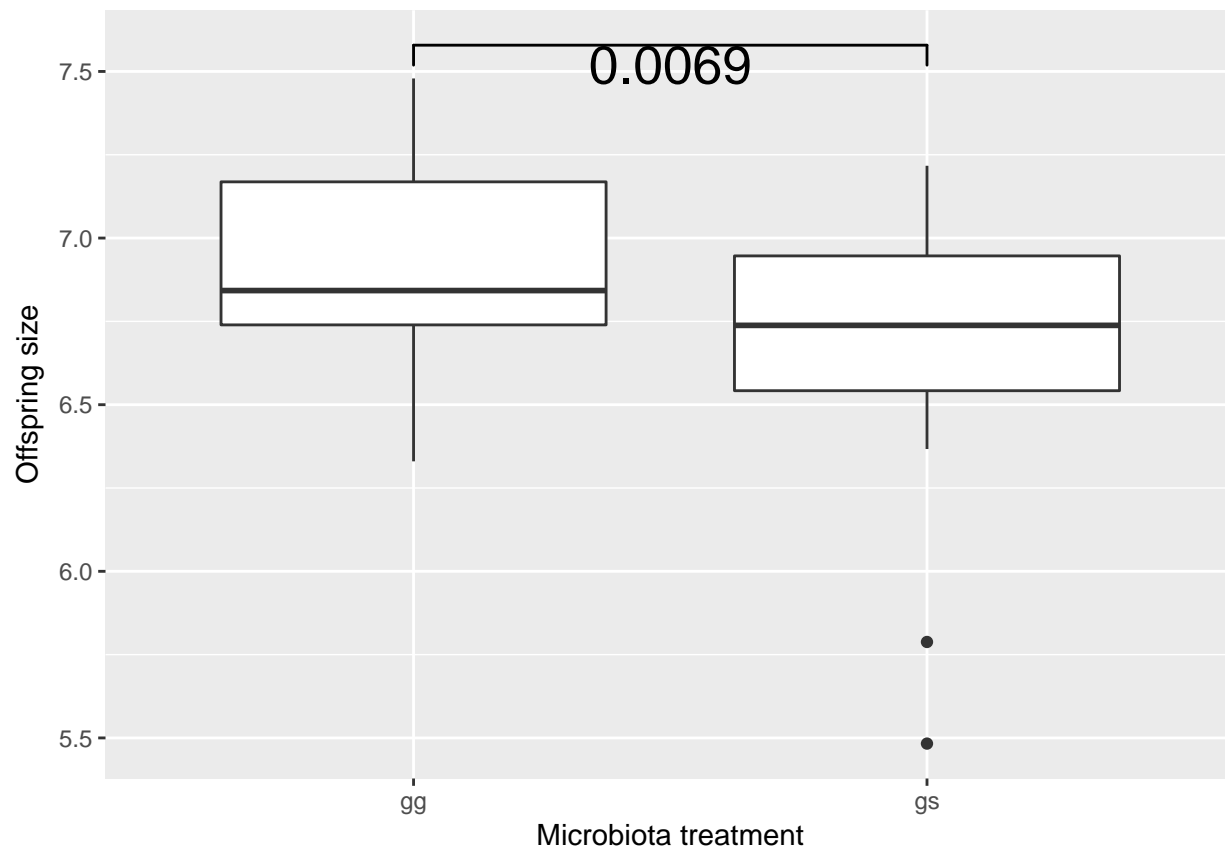


First, though there are unfortunately not many mothers in this analysis, we see that we can still recover the normal trend of increasing maternal size -> increasing offspring size as we expect. As we knew going in that we didn't have a lot of mothers alive for this part of the study, we measured individual brood ball sizes of the F2s because this is generally taken as the primary way that mothers invest in their offspring. Doing this, we were still able to control for maternal effects - even without a lot of variation in maternal size.

```
ggplot(f2.validate, aes(x = Treatment.ID, y = BB.weight)) + geom_boxplot() +  
  geom_signif(comparisons = list(c("gg", "gs")), test = "wilcox.test", map_signif_level = FALSE,  
  textsize = 7, vjust = +1.2) + labs(y = "BB weight", x = "Microbiota treatment")
```



```
ggplot(f2.validate, aes(x = Treatment.ID, y = Scope.size)) + geom_boxplot() +
  geom_signif(comparisons = list(c("gg", "gs")), test = "wilcox.test", map_signif_level = FALSE,
    textsize = 7, vjust = +1.2) + labs(y = "Offspring size", x = "Microbiota treatment")
```



Also, we do see a significant difference (from a Wilcoxon signed-rank test) in the brood ball sizes produced by the mothers who received different treatments early in development, and also a significant difference in offspring size of mothers from different treatment groups. This together suggests to me that there is likely an interaction between our proxy for a maternal effect, brood ball size, and treatment.

```
m3 <- lm(Scope.size ~ Sex + Treatment.ID + BB.weight + Treatment.ID:BB.weight +
  Sex:Treatment.ID, data = f2.validate)
```

```
Anova(m3)
```

```
## Anova Table (Type II tests)
```

```
##
```

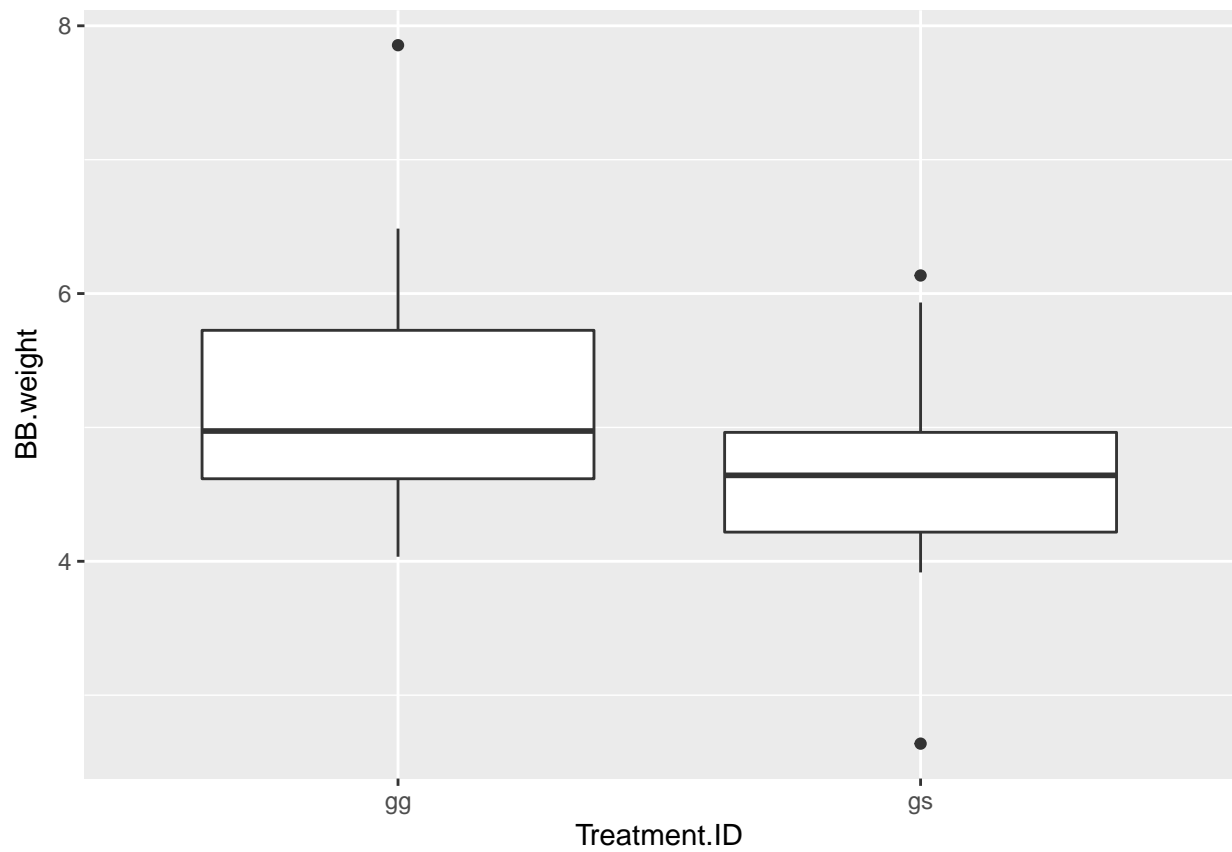
```
## Response: Scope.size
```

	Sum Sq	Df	F value	Pr(>F)
## Sex	1.9460	1	38.3269	3.758e-08 ***
## Treatment.ID	0.9203	1	18.1248	6.404e-05 ***
## BB.weight	0.0143	1	0.2822	0.5970
## Treatment.ID:BB.weight	0.9592	1	18.8916	4.665e-05 ***
## Sex:Treatment.ID	0.2251	1	4.4325	0.0389 *
## Residuals	3.5034	69		

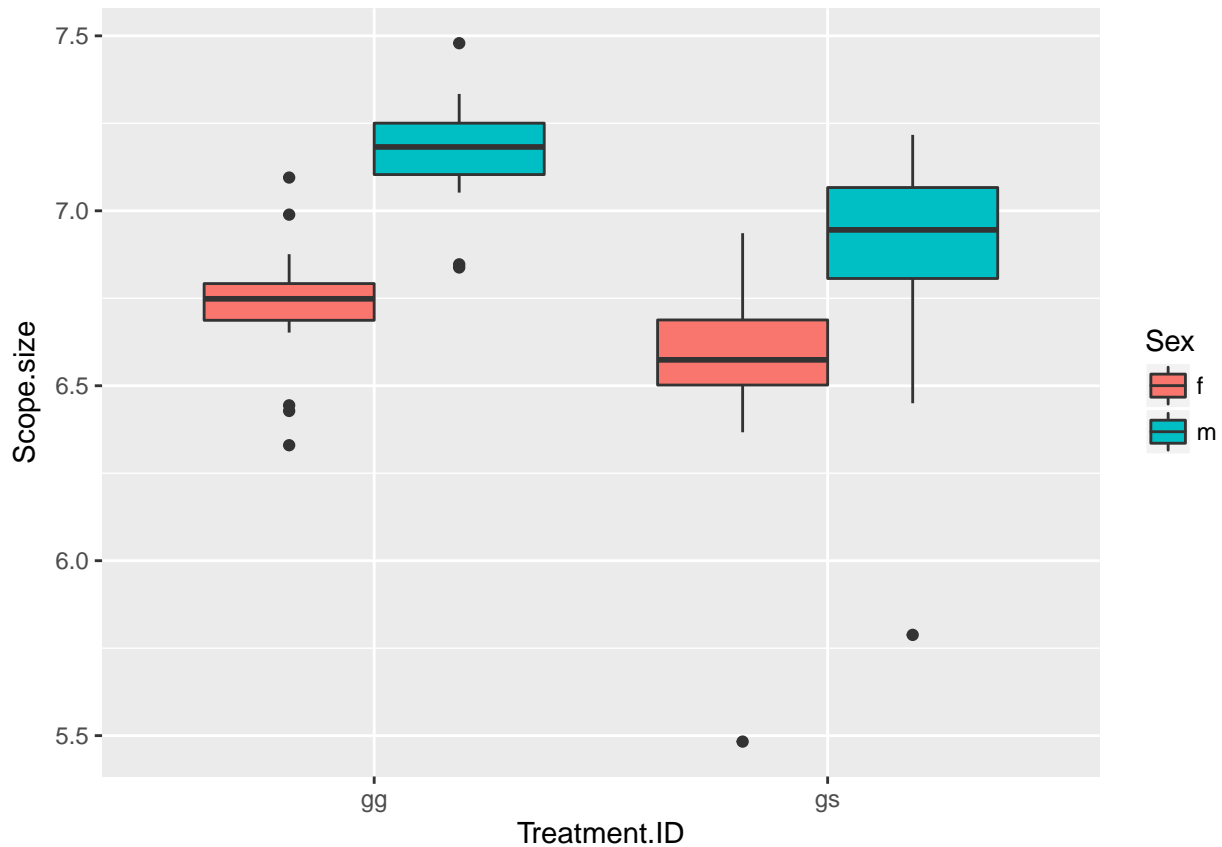
```
## ---
```

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

```
ggplot(f2.validate, aes(x = Treatment.ID, y = BB.weight)) + geom_boxplot()
```



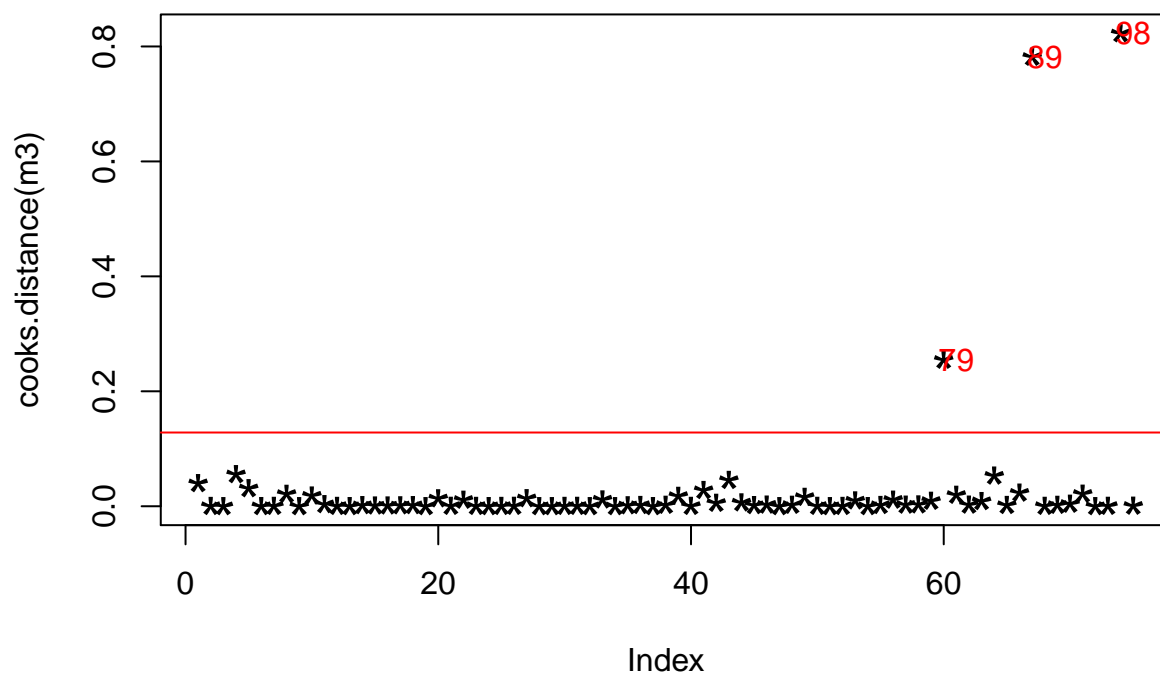
```
ggplot(f2.validate, aes(x = Treatment.ID, y = Scope.size, fill = Sex)) + geom_boxplot()
```



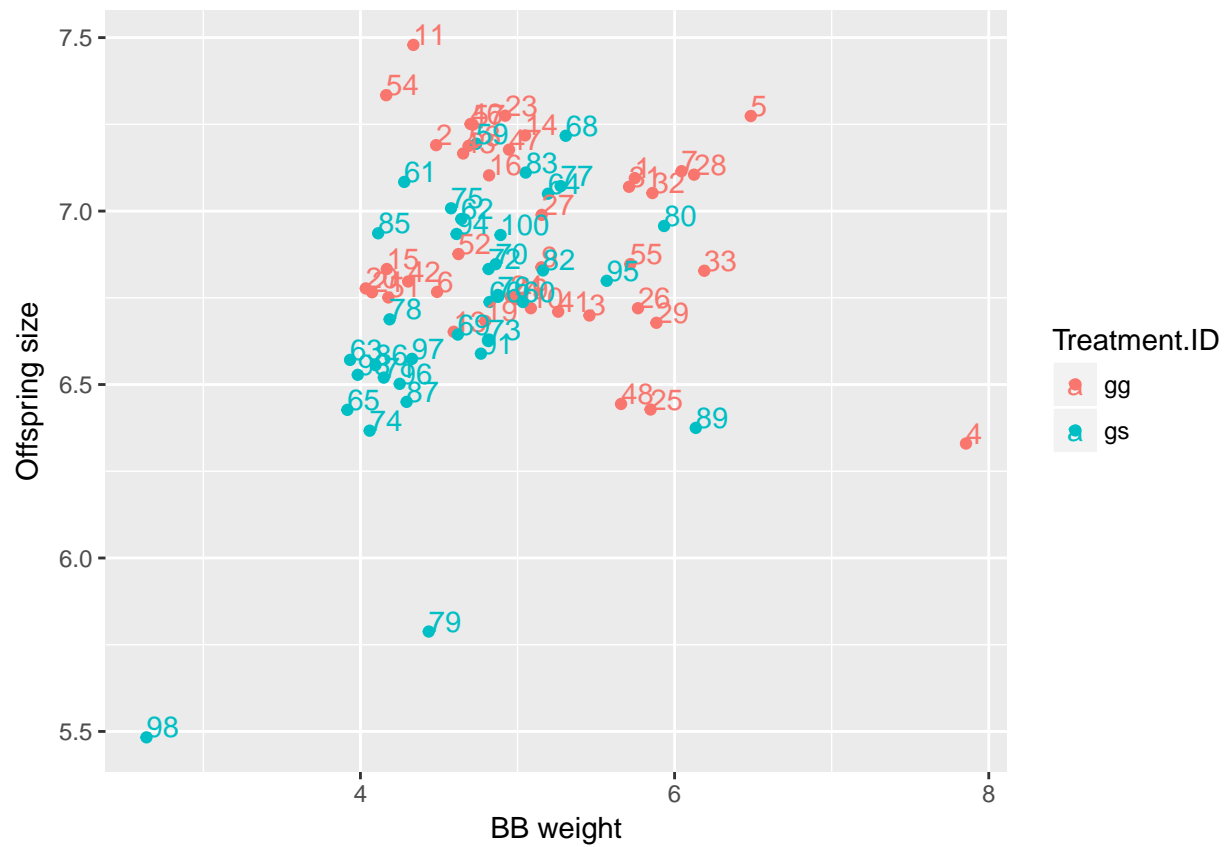
Through model selection, I find that the best model for the raw data is one which contains an interaction between brood ball size and treatment (as well as one between sex and treatment to a lesser extent). This is not really suprising given the plots shown above. However, when we check this model for outlying and influential points, there is reason to be concerned with the conclusions of this model.

```
plot(cooks.distance(m3), pch = "*", cex = 2, main = "Influential Obs by Cooks distance")
abline(h = 4 * mean(cooks.distance(m3), na.rm = T), col = "red")
text(x = 1:length(cooks.distance(m3)) + 1, y = cooks.distance(m3), labels = ifelse(cooks.distance(m3) >
  4 * mean(cooks.distance(m3), na.rm = T), names(cooks.distance(m3)), ""),
  col = "red")
```

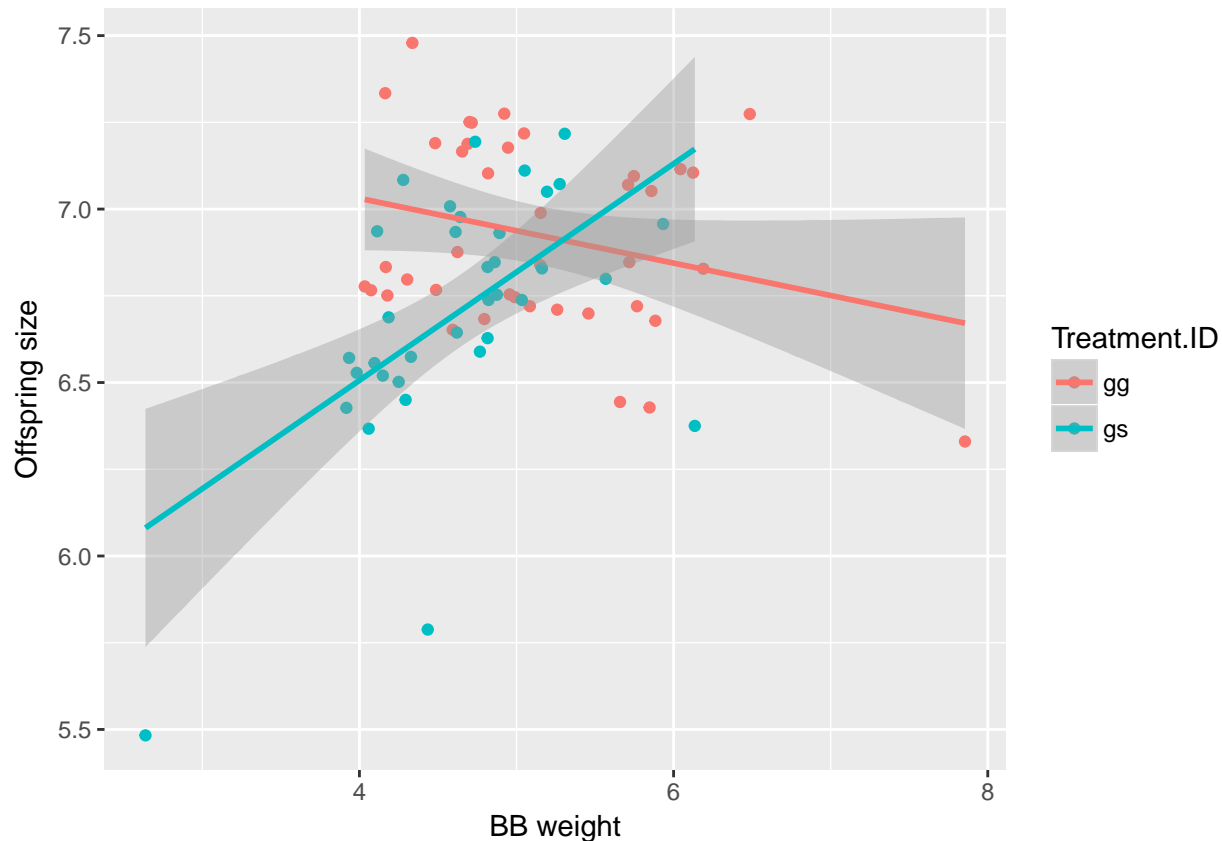
Influential Obs by Cooks distance



```
ggplot(f2.validate, aes(x = BB.weight, y = Scope.size, color = Treatment.ID,
  label = rownames(f2.validate))) + geom_point() + geom_text(aes(label = rownames(f2.validate)),
  hjust = 0, vjust = 0) + labs(y = "Offspring size", x = "BB weight")
```



```
ggplot(f2.validate, aes(x = BB.weight, y = Scope.size, color = Treatment.ID)) +
  geom_point() + labs(y = "Offspring size", x = "BB weight") + geom_smooth(method = "lm")
```



First, we see that a few datapoints are highly influential by cook's distance. The 3 most influential seem to be observations 79, 89, and 98. Then, from an analysis of the plot of offspring size vs. bb weight, we see why. Observations 98 and 79 are quite clearly outliers - though 89 doesn't seem too bad, it is much closer to the rest of the distribution. Furthermore it's not as clearcut but observation 4, which was not identified as highly influential by cook's distance, could also be considered an outlier.

The worst points are definitely observations 98 and 79 though, it is clear to see that they are likely largely responsible for the significant interaction between bb size and treatment identified in the previous model. I thus decided to remove them and go through another round of model selection to test this theory.

The outcome of this was basically that depending on which of these potential outliers I decided to remove, the interaction between bb size and treatment was either found to be significant or not. If I remove observations 98 and 79 (the clearest outliers), the interaction is still "significant", but just barely. But if I additionally remove observation 4 which is *extremely* borderline, and I keep going back and forth on the merits of including it, the interaction between bb size and treatment is no longer significant, and the model is just reduced to sex and treatment (what is currently presented in the paper). What this suggests to me is that there is *very likely* a maternal effect here in reality, it just isn't captured convincingly in these data due to the small sample size.

All that said, I'm still not completely sure how comfortable I am leaving observation 4 in the data (and thus including the maternal effect), as it alone causes the conspecific treatment (gg) to show a relationship of decreasing offspring weight with increasing bb size - which is almost surely not reflective of reality. This is the main reason I chose to also exclude point 4 from the currently presented analysis.

```
f2.outliers <- f2.validate[-c(60, 74), ]
# removing points 79 and 98.
```



```

m1 <- lm(Scope.size ~ Sex + Treatment.ID + BB.weight + Treatment.ID:BB.weight +
  Sex:Treatment.ID, data = f2.outliers)

m2 <- lm(Scope.size ~ Sex + Treatment.ID, data = f2.outliers)

m3 <- lm(Scope.size ~ Sex + Treatment.ID + BB.weight + BB.weight:Treatment.ID,
  data = f2.outliers)

anova(m2, m1)

## Analysis of Variance Table
##
## Model 1: Scope.size ~ Sex + Treatment.ID
## Model 2: Scope.size ~ Sex + Treatment.ID + BB.weight + Treatment.ID:BB.weight +
##   Sex:Treatment.ID
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      70 2.1226
## 2      67 1.8165  3   0.30606 3.7628 0.0147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# reject reduced model m2 - close call though.

anova(m3, m1)

## Analysis of Variance Table
##
## Model 1: Scope.size ~ Sex + Treatment.ID + BB.weight + BB.weight:Treatment.ID
## Model 2: Scope.size ~ Sex + Treatment.ID + BB.weight + Treatment.ID:BB.weight +
##   Sex:Treatment.ID
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      68 1.8935
## 2      67 1.8165  1   0.076948 2.8381 0.09671 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# Also a close call, but accept reduced model.

Anova(m3)

## Anova Table (Type II tests)
##
## Response: Scope.size
##           Sum Sq Df F value    Pr(>F)
## Sex          2.41461  1 86.7150 9.253e-14 ***
## Treatment.ID    0.68015  1 24.4260 5.315e-06 ***
## BB.weight       0.11077  1  3.9781  0.05011 .
## Treatment.ID:BB.weight 0.11834  1  4.2498  0.04308 *
## Residuals      1.89348 68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

calc.relimp(m3, type = "lmg")

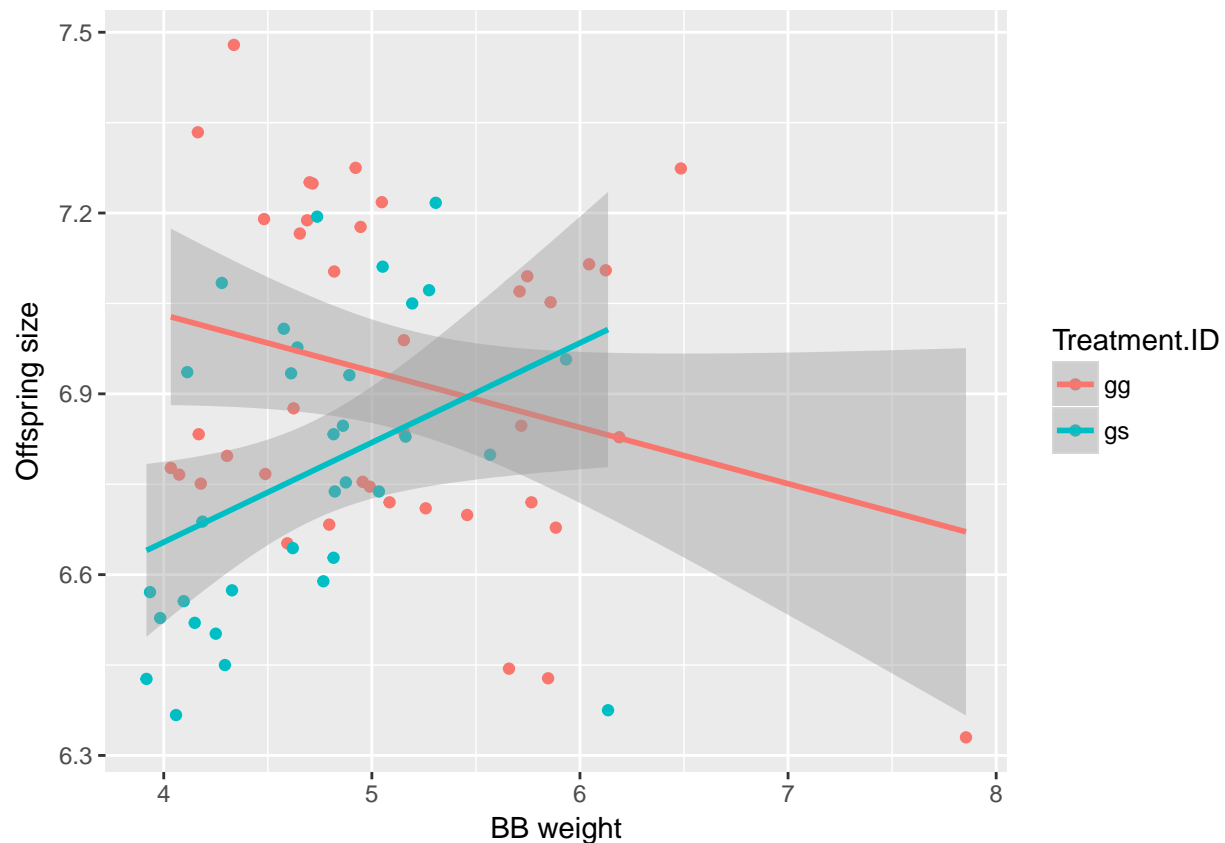
## Response variable: Scope.size
## Total response variance: 0.07253941

```

```
## Analysis based on 73 observations
##
## 4 Regressors:
## Sex Treatment.ID BB.weight Treatment.ID:BB.weight
## Proportion of variance explained by model: 63.75%
## Metrics are not normalized (rela=FALSE).
##
## Relative importance metrics:
##
##                                lmg
## Sex                        0.493379850
## Treatment.ID              0.097715384
## BB.weight                 0.006645607
## Treatment.ID:BB.weight 0.039720280
##
## Average coefficients for different model sizes:
##
##                                1X          2Xs          3Xs          4Xs
## Sex                        0.37222932  0.3790444  0.39441112  0.3767318
## Treatment.ID              -0.15383485 -0.1704076 -0.80798178 -0.8408673
## BB.weight                 0.01497132 -0.0159146 -0.07524407 -0.0935676
## Treatment.ID:BB.weight      NaN          NaN  0.25861340  0.1321105

# Even when bb weight and bbweight:treatment is included, treatment effect
# by itself explains ~10% of variance still. About the same amount as in the
# reduced model with just sex and treatment (shown below).

ggplot(f2.outliers, aes(x = BB.weight, y = Scope.size, color = Treatment.ID)) +
  geom_point() + labs(y = "Offspring size", x = "BB weight") + geom_smooth(method = "lm")
```



Case where observation 4 is left, plot shows an unrealistic seeming relationship for gg treatment because of this one point.

```
f2.outliers <- f2.validate[-c(4, 74, 60), ]
# removing points 4, 79, and 98.

m1 <- lm(Scope.size ~ Sex + Treatment.ID + BB.weight + Treatment.ID:BB.weight +
  Sex:Treatment.ID, data = f2.outliers)

m2 <- lm(Scope.size ~ Sex + Treatment.ID, data = f2.outliers)

m3 <- lm(Scope.size ~ Sex + Treatment.ID + BB.weight + BB.weight:Treatment.ID,
  data = f2.outliers)

anova(m2, m1)

## Analysis of Variance Table
##
## Model 1: Scope.size ~ Sex + Treatment.ID
## Model 2: Scope.size ~ Sex + Treatment.ID + BB.weight + Treatment.ID:BB.weight +
##       Sex:Treatment.ID
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      69 1.9381
## 2      66 1.7843  3    0.1538 1.8963 0.1388

anova(m2, m3)

## Analysis of Variance Table
```

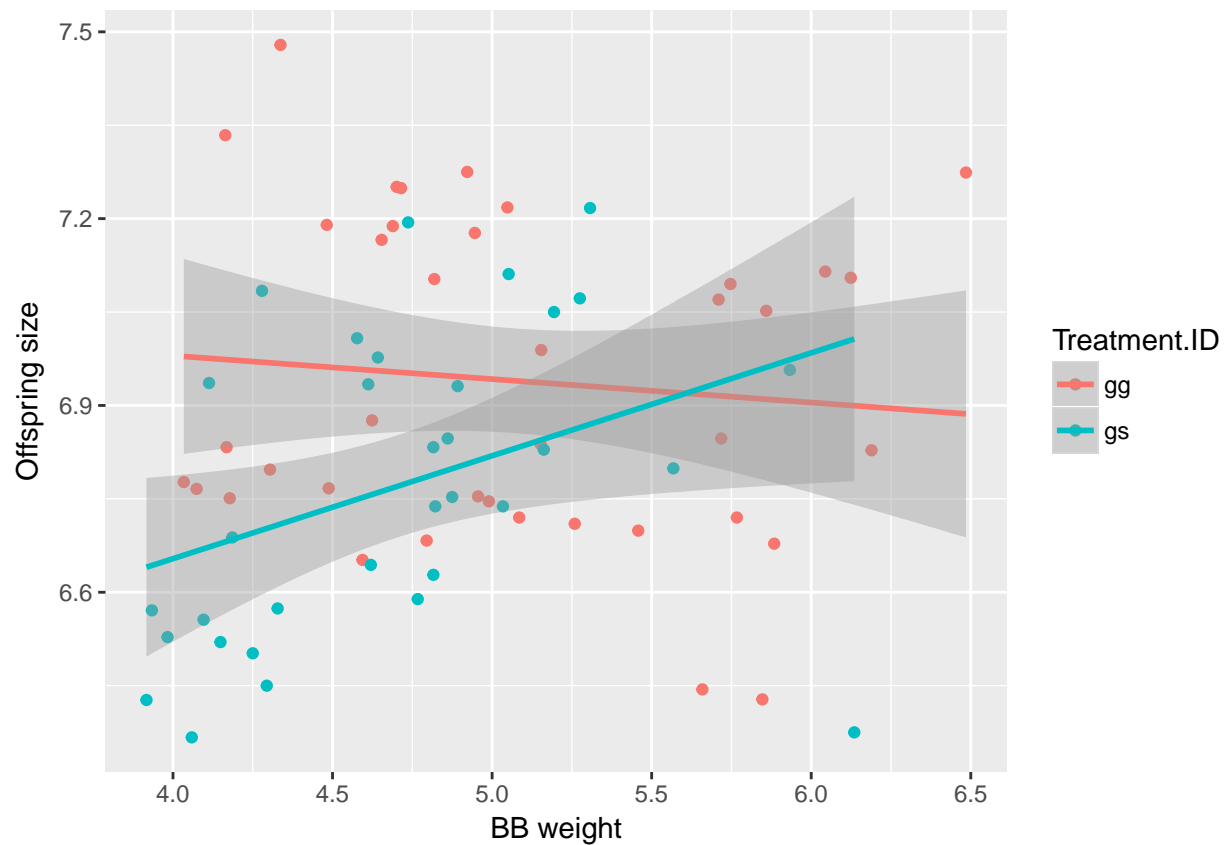
```
##
## Model 1: Scope.size ~ Sex + Treatment.ID
## Model 2: Scope.size ~ Sex + Treatment.ID + BB.weight + BB.weight:Treatment.ID
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      69 1.9381
## 2      67 1.8494  2  0.088705 1.6068 0.2082

# Can't reject null hypothesis that reduced model explains the data well
# enough.

calc.relimp(m2, type = "lmg")

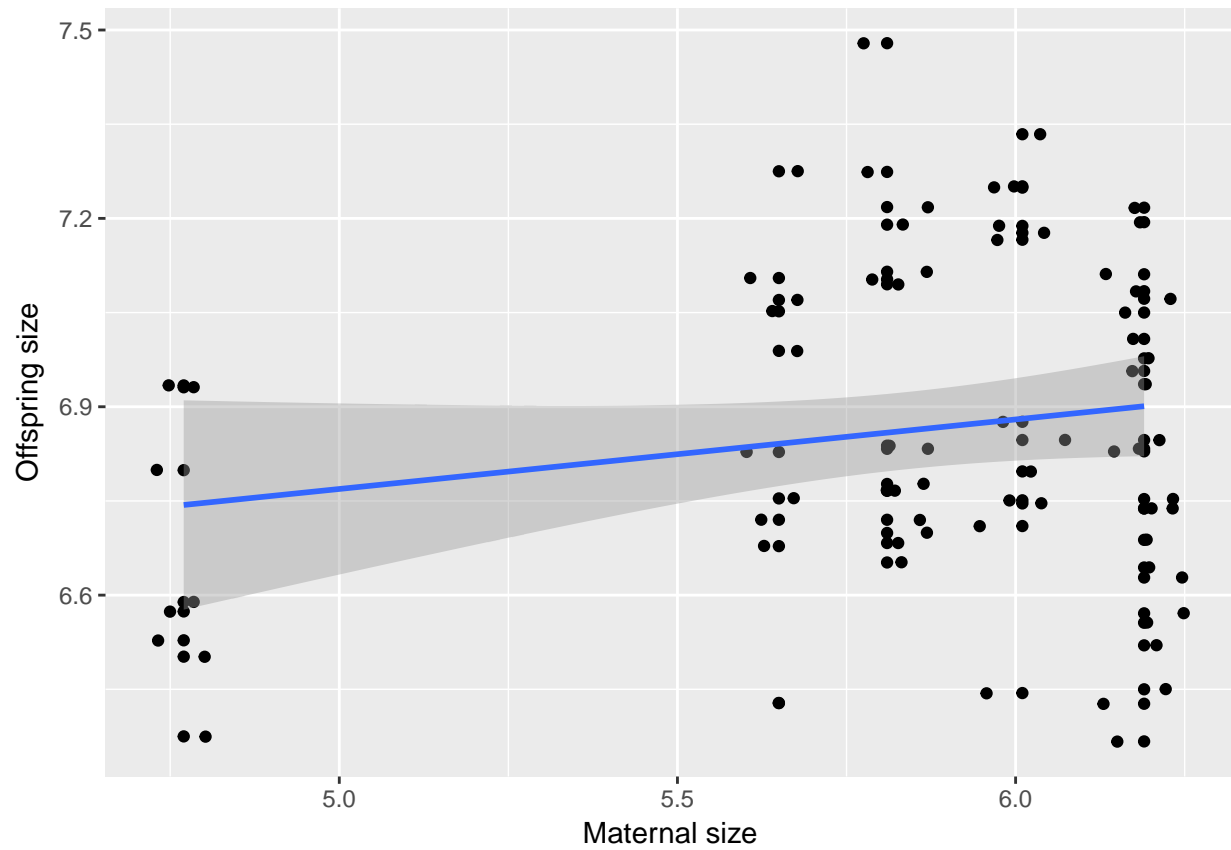
## Response variable: Scope.size
## Total response variance: 0.069628
## Analysis based on 72 observations
##
## 2 Regressors:
## Sex Treatment.ID
## Proportion of variance explained by model: 60.79%
## Metrics are not normalized (rela=FALSE).
##
## Relative importance metrics:
##
##               lmg
## Sex           0.4918262
## Treatment.ID 0.1161235
##
## Average coefficients for different model sizes:
##
##           1X           2Xs
## Sex       0.3628687 0.3729427
## Treatment.ID -0.1690746 -0.1890692

ggplot(f2.outliers, aes(x = BB.weight, y = Scope.size, color = Treatment.ID)) +
  geom_point() + labs(y = "Offspring size", x = "BB weight") + geom_smooth(method = "lm")
```



Now when observation 4 is taken out, the interaction that was significant before no longer is and the maternal effect part of the model drops out. But to me this looks like a correct decision because the plot looks much better.

```
ggplot(f2.outliers, aes(x = Maternal.size, y = Scope.size)) + geom_point() +
  geom_jitter() + geom_smooth(method = "lm") + labs(y = "Offspring size",
x = "Maternal size")
```



But, even when all 3 of the most suspicious points are dropped, still see the trend we expect (which was also reflected in the raw data) - larger moms generally have larger offspring.