Plasticity in Littorina offspring, the size issue

Samuel

4/2/2020

- bold and PC2 seem to be size-independent. That is good!

- thickness is clearly size-dependent. There is some scatter, which might mean that there is some interesting size-independent thickness variation but it might also just mean that the thickness measurements are noisy!

- cube\_root\_weight looks almost perfectly correlated with size! Also, the relationship looks linear across the whole size range and very similar in all populations. t may be that there is no size-independent variation in weight.

For the weight along the cline plots:

- Gen0 is the expected pattern, which is good.

* Gen1 is flat, as your said. This is a bit odd.

There are two obvious alternative explanations (not mutually exclusive): sampling was timed for when the size reached a certain point, or size is very plastic. We might separate these effects (partly) by correcting for sampling time (although larger snails were taken at each time point, I guess). Ultimately, this sampling time issue might make it hard to say anything about plasticity for size (or weight).

And the stats:

- The linear model needs to include the interaction term so that we can see if there is any variation in the slope of the regression among tanks (or positions in the transect). Actually, we need to test whether including tank and tank.size improve the model, compared to size only.

- tank C, gen0 has a strong outlier, which seems to be influencing the results. Can you correct or remove?

For now, the data and results belong to the island CZA.

## Part one: Size dependence of other traits

A plastic response to a change in the environment can be observed in one or more phenotypes. Ideally, we would like to understand to what extent each phenotype is plastic but in some (perhaps many) cases the plastic responses are not independent because the phenotypic values are correlated.  
Our question is about plasticity which can be studied from the relationship between offspring and parents (by tank). We think that any snail will change in various phenotypes as it grows. Taking weight as an example, the bigger the snail, the heavier it is. This means that comparing weights of offspring and parents risks simply showing us effects of size, rather than the plasticity in weight that we are interested in. We might see effects of size plasticity but we might also see confounding effects of other types of size variation, particularly the effect of sampling time.  
There are generally two ways to deal with this: find the relationship between weight and size and then analyse residuals from that relationship, or include size as a covariate in the model when analysing weight. Generally, the latter approach is preferable but it can be hard to implement correctly (and I think that is probably true in our case because of the unusual regression approach we are using). Both approaches have a problem for the snails: we expect that (at least for some traits) the effect of size will be different for crab and wave (and might change continuously across the contact zone). So, there may be no single relationship between size and weight (for example) that can be applied to generate residuals.  
I think this all means that the first step is to test the relationship of each variable to size and test whether the relationship is constant over tanks. I think this may have to be done separately for parents and offspring (although it might be nice to test whether the relationship changes…) and maybe also for adults and juveniles.  
For weight, we expect and weight should scale with , where size is measured as length (mm) so an alternative is to use . So, we should fit:

(and try adding a term for maturity?).

## [[1]]

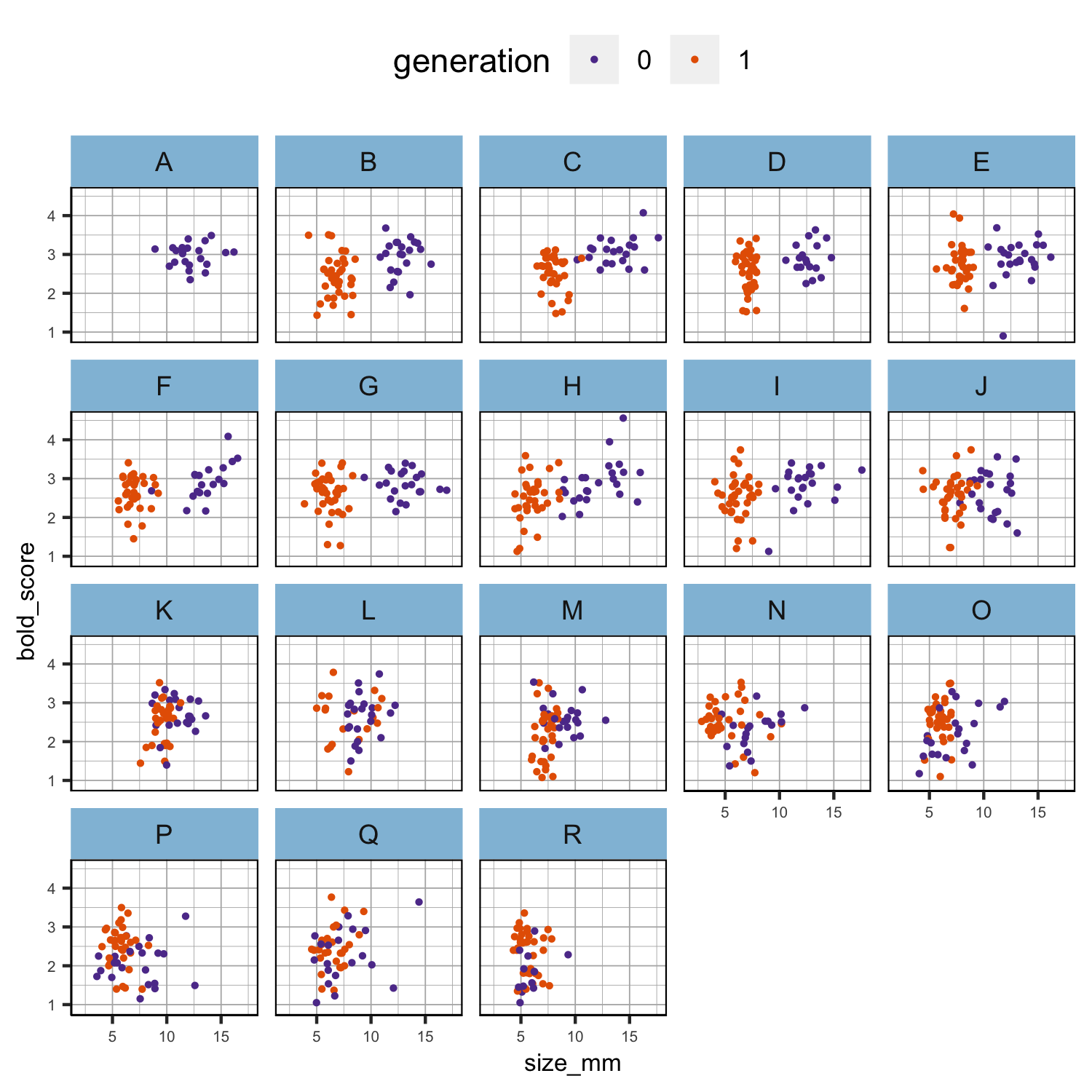


Figure tmp1. Scatter plots of the relationship between size and the other phenotypes per population in each generation separately. Phenotypic values have not been scaled.

##   
## [[2]]

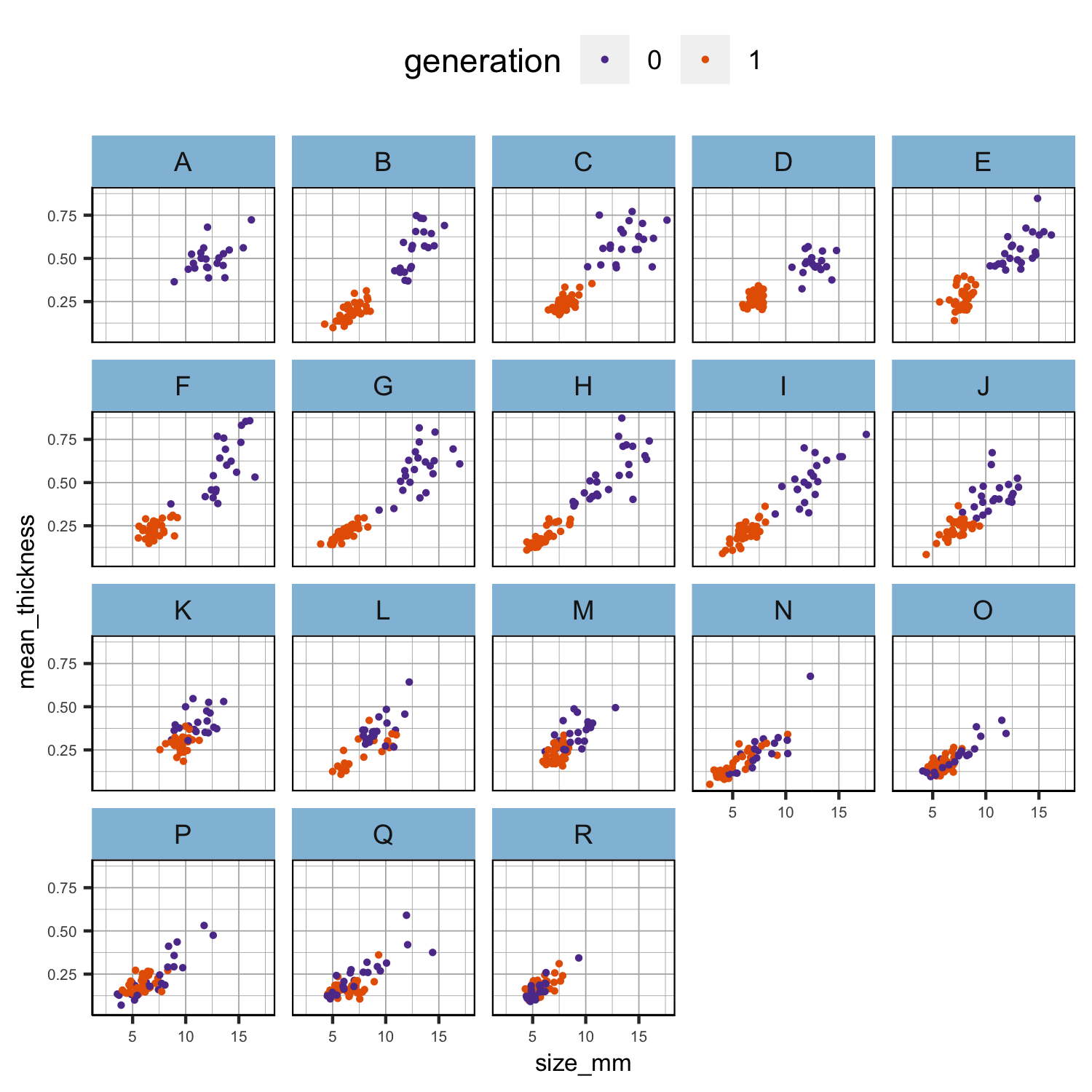


Figure tmp1. Scatter plots of the relationship between size and the other phenotypes per population in each generation separately. Phenotypic values have not been scaled.

##   
## [[3]]

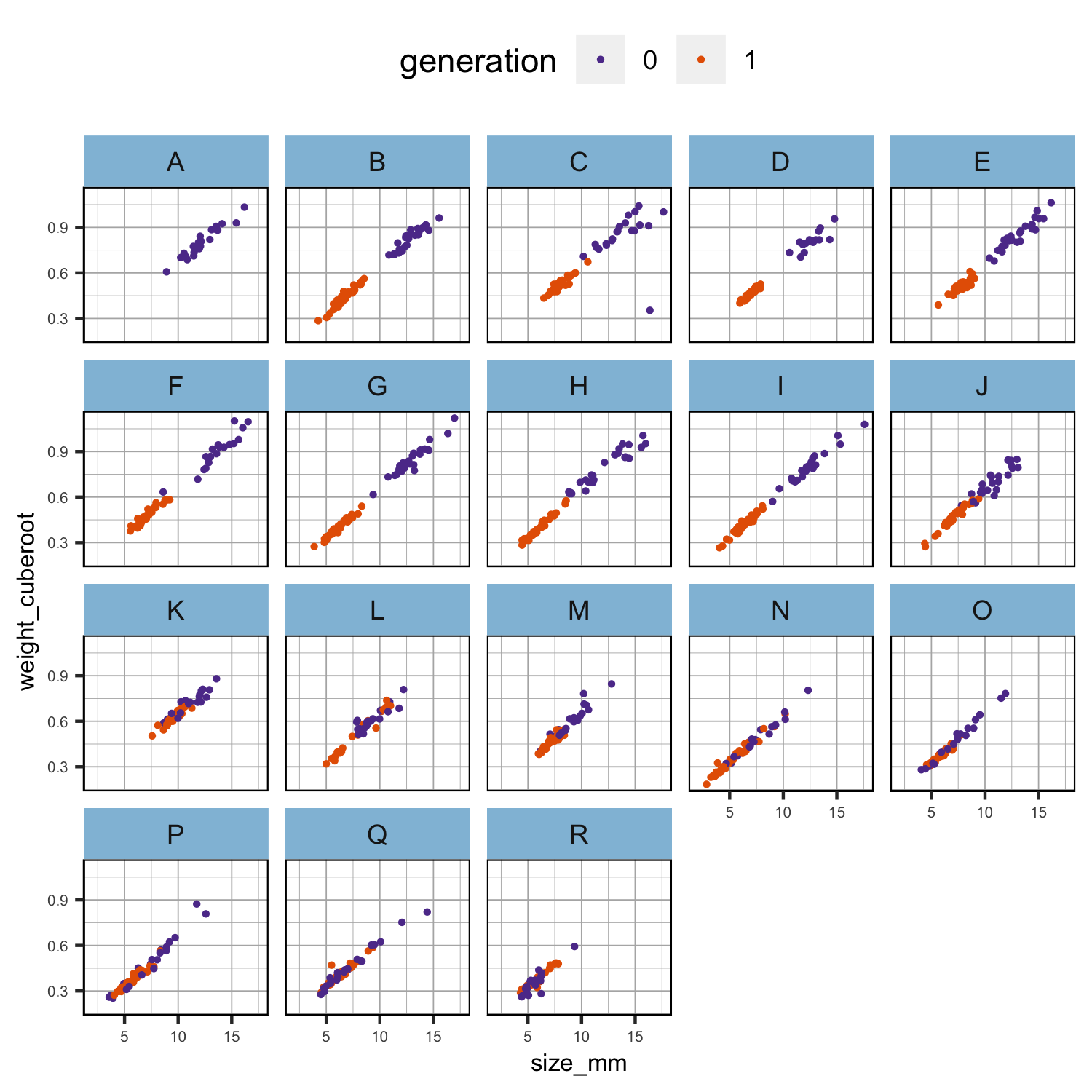


Figure tmp1. Scatter plots of the relationship between size and the other phenotypes per population in each generation separately. Phenotypic values have not been scaled.

##   
## [[4]]

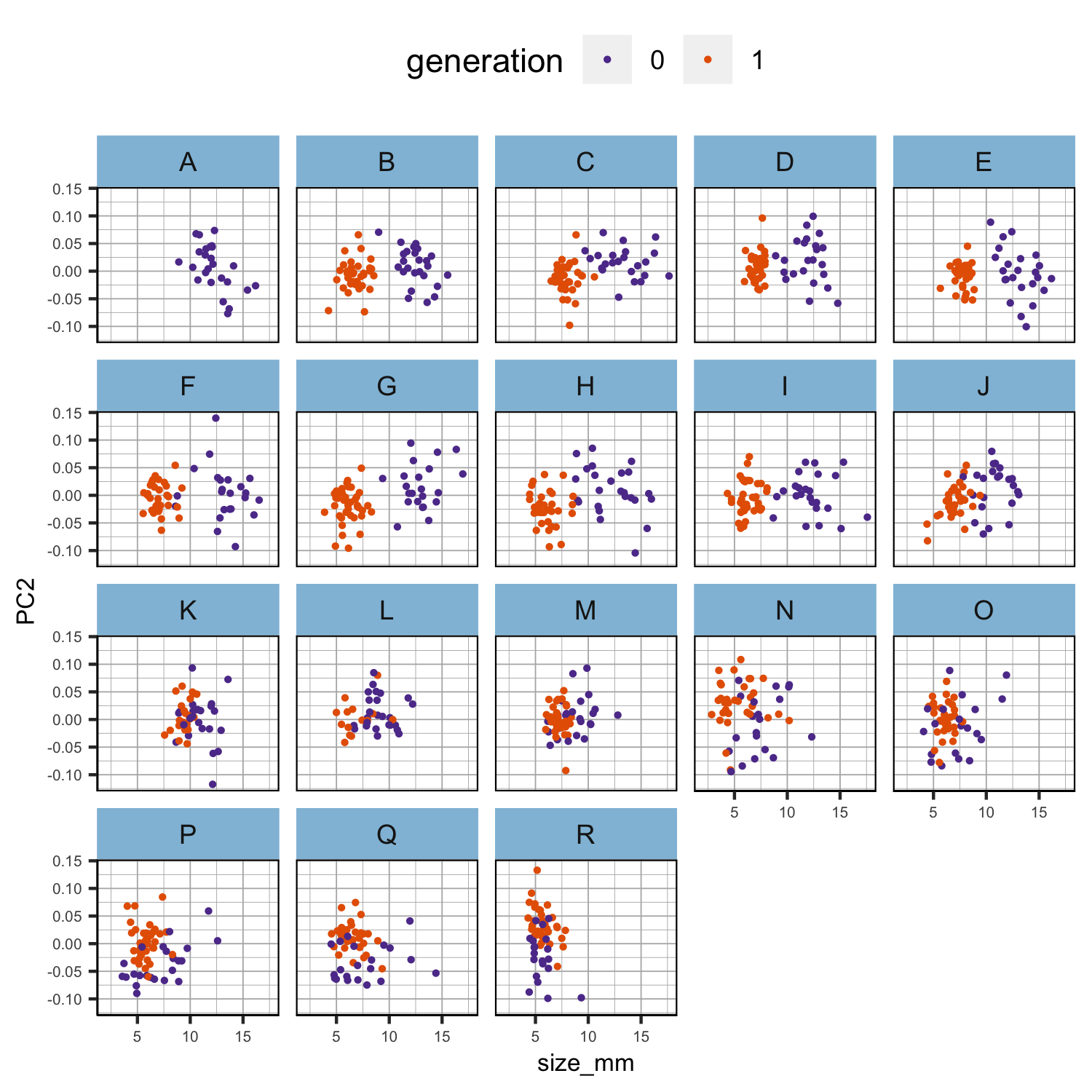
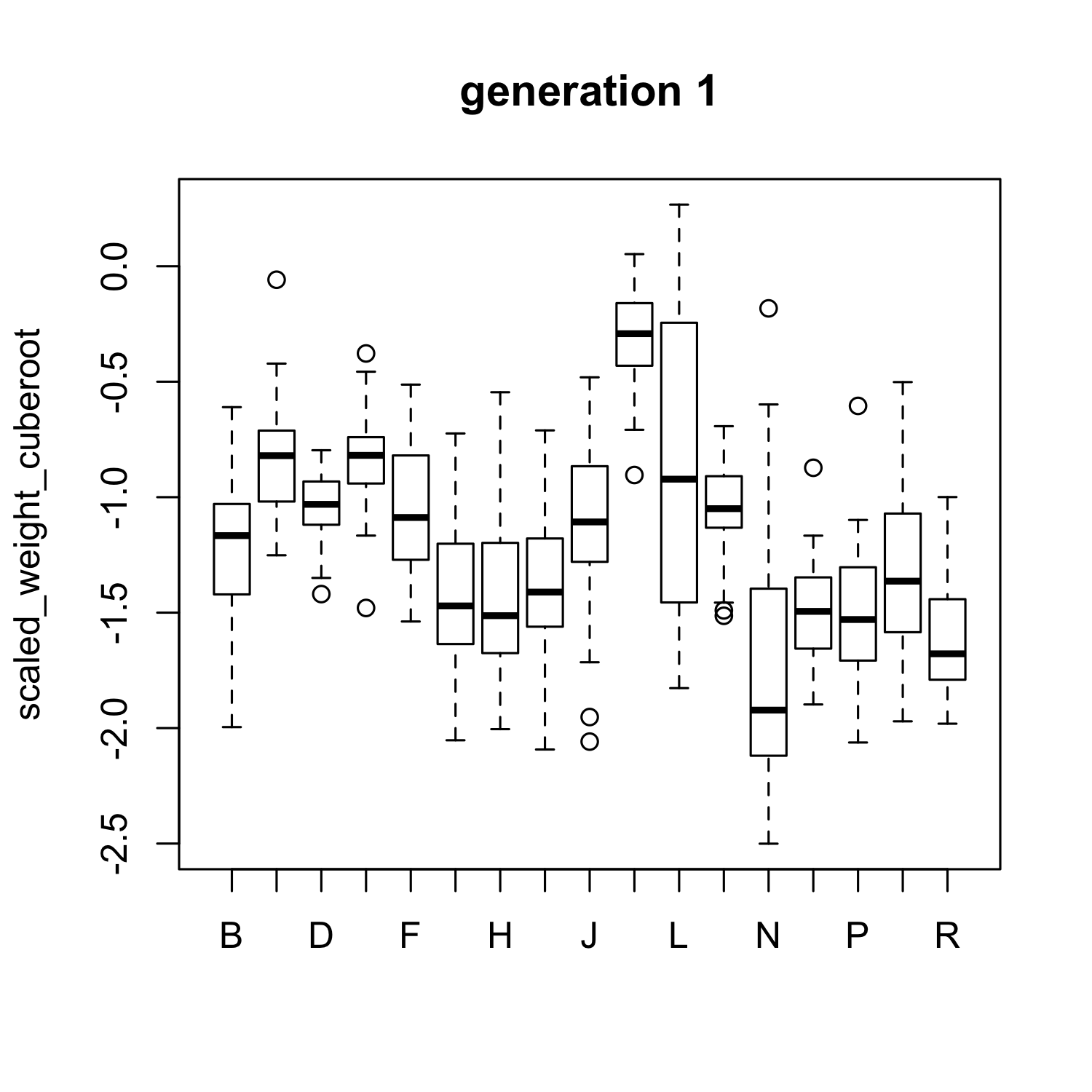
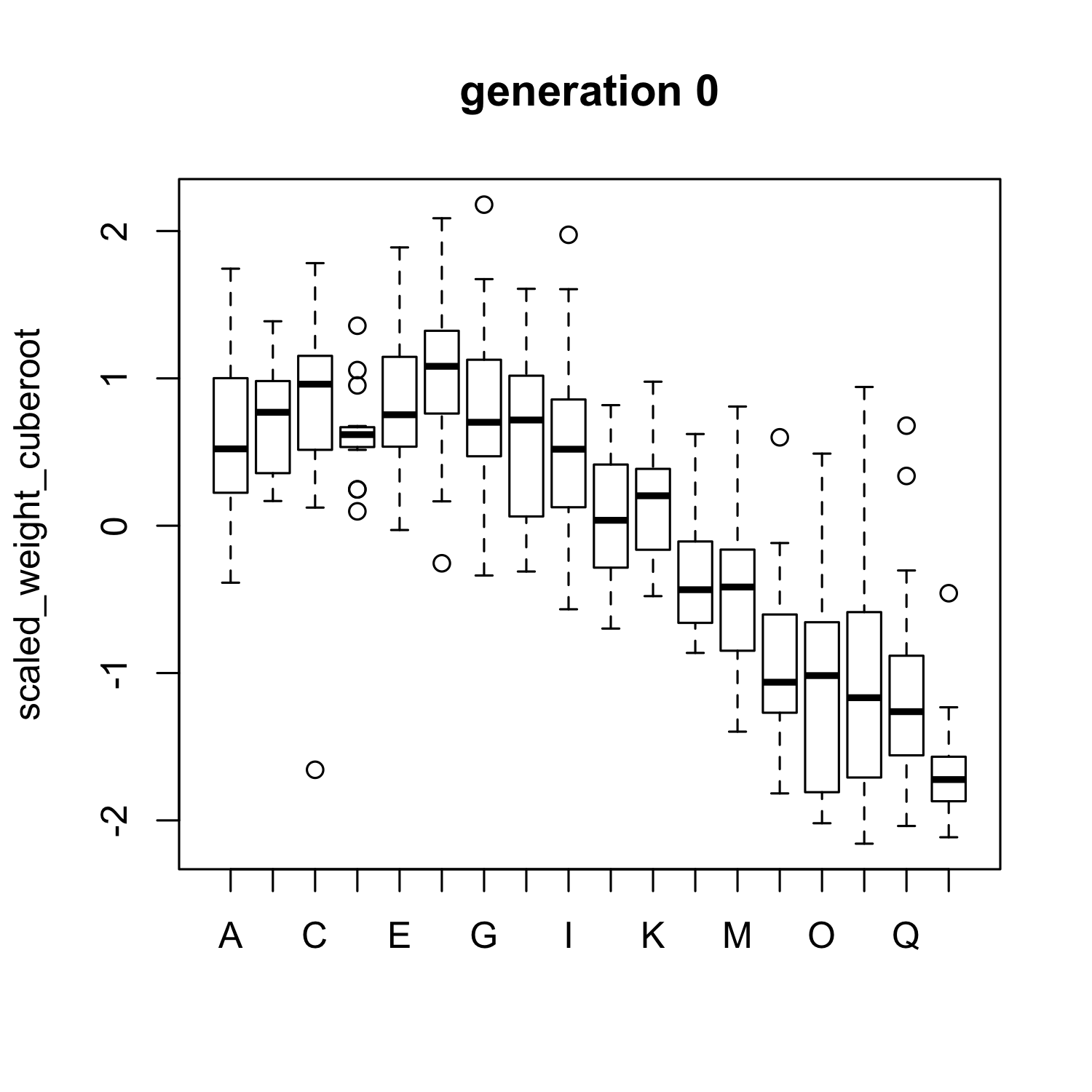


Figure tmp1. Scatter plots of the relationship between size and the other phenotypes per population in each generation separately. Phenotypic values have not been scaled.



*Figure tmp2. Boxplot of the scaled cube root of weight per generation and population. Phenotypic values have been scaled.*

Table tmp1a. Summary of the linear model in generation 0 parents with scaled cube root of the weight as response variable and scaled size plus population as explanatory variables.

| Parameter | Estimate | Std..Error | t.value | Pr...t.. |
| --- | --- | --- | --- | --- |
| (Intercept) | 0.104 | 0.053 | 1.975 | 0.049 |
| scaled\_size\_mm | 0.880 | 0.022 | 40.602 | 0.000 |
| popB | -0.047 | 0.072 | -0.646 | 0.518 |
| popC | -0.209 | 0.074 | -2.824 | 0.005 |
| popD | -0.055 | 0.076 | -0.726 | 0.469 |
| popE | 0.002 | 0.071 | 0.029 | 0.977 |
| popF | 0.093 | 0.075 | 1.249 | 0.213 |
| popG | -0.002 | 0.072 | -0.025 | 0.980 |
| popH | -0.009 | 0.070 | -0.135 | 0.893 |
| popI | -0.060 | 0.073 | -0.817 | 0.415 |
| popJ | -0.098 | 0.072 | -1.356 | 0.176 |
| popK | -0.084 | 0.073 | -1.153 | 0.249 |
| popL | -0.112 | 0.075 | -1.484 | 0.139 |
| popM | -0.085 | 0.076 | -1.120 | 0.263 |
| popN | -0.239 | 0.083 | -2.893 | 0.004 |
| popO | -0.231 | 0.081 | -2.845 | 0.005 |
| popP | -0.193 | 0.080 | -2.410 | 0.016 |
| popQ | -0.291 | 0.080 | -3.635 | 0.000 |
| popR | -0.365 | 0.087 | -4.211 | 0.000 |

Table tmp1b. Summary of the linear model in generation 1 offspring with scaled cube root of the weight as response variable and scaled size plus population as explanatory variables.

| Parameter | Estimate | Std..Error | t.value | Pr...t.. |
| --- | --- | --- | --- | --- |
| (Intercept) | -0.061 | 0.019 | -3.225 | 0.001 |
| scaled\_size\_mm | 0.951 | 0.010 | 90.718 | 0.000 |
| popC | 0.025 | 0.020 | 1.220 | 0.223 |
| popD | 0.057 | 0.020 | 2.875 | 0.004 |
| popE | 0.043 | 0.020 | 2.101 | 0.036 |
| popF | 0.078 | 0.020 | 3.894 | 0.000 |
| popG | -0.027 | 0.019 | -1.395 | 0.163 |
| popH | 0.008 | 0.020 | 0.404 | 0.687 |
| popI | 0.015 | 0.020 | 0.755 | 0.451 |
| popJ | 0.022 | 0.020 | 1.092 | 0.275 |
| popK | 0.043 | 0.023 | 1.865 | 0.063 |
| popL | 0.002 | 0.024 | 0.094 | 0.925 |
| popM | 0.000 | 0.020 | 0.001 | 0.999 |
| popN | -0.010 | 0.020 | -0.466 | 0.641 |
| popO | -0.063 | 0.020 | -3.130 | 0.002 |
| popP | -0.018 | 0.020 | -0.926 | 0.355 |
| popQ | -0.030 | 0.020 | -1.494 | 0.136 |
| popR | -0.047 | 0.020 | -2.360 | 0.019 |

It may then be possible to predict what the mean for snails in each tank would be at the overall mean size of snails for the experiment (with associated standard error) and plug that into the plasticity calculations.

Generally, I do not think this is a multi-colinearity problem because we are not trying to explain one response variable on the basis of many independent variables. It is more productive to think of each variable as a function of size. I know that size + thickness might completely explain weight but that is a somewhat separate issue. It would just make plasticity in weight rather uninteresting because it would depend on size and thickness plasticity.

### Results

As expected, size, thickness and weight result to be highly correlated within generations (Fig. 1) and between generations (Fig. 2). One way to detect whether they are also collinear is to check the Choleski decomposition of the correlation matrix - if there is (multi)collinearity there will be some diagonal elements that are close to zero (Table 2). There is also another way to assess multicollinearity and that is by computing a score called the variance inflation factor (VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model. I have not exactly figured out how to implement this second test for multicollinearity to our case but, for example, if we consider the outcome variable to be the thickness of the offspring, then we can build a regression model with four predictors, thickness, weight, shape (PC2) and size of the parents, and calculate the VIF score after model fitting. Does it make sense?

## [[1]]

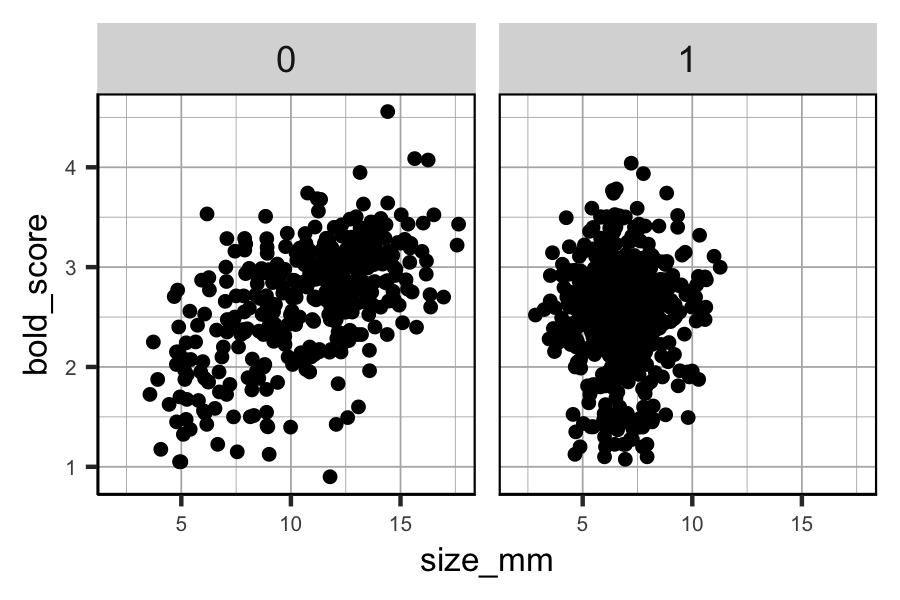


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[2]]

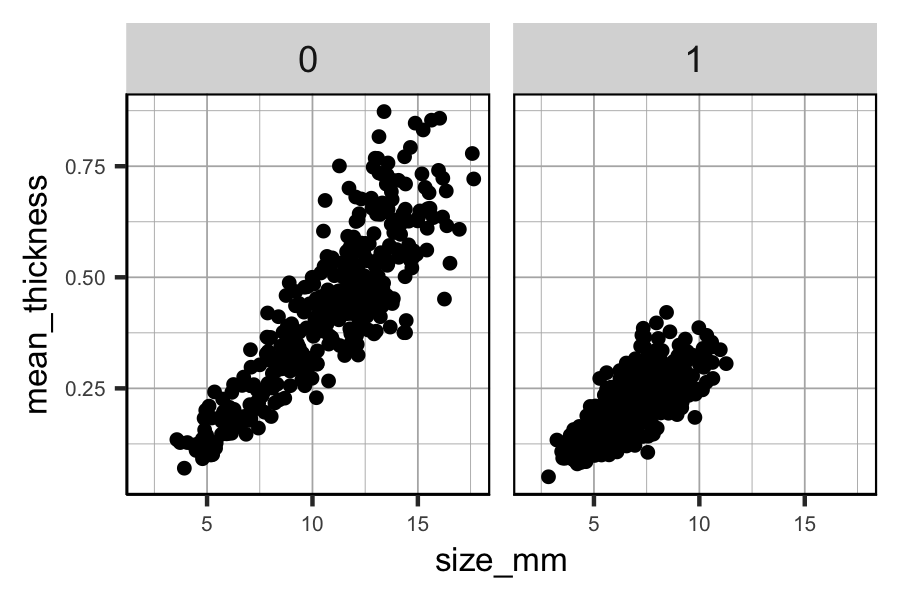


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[3]]

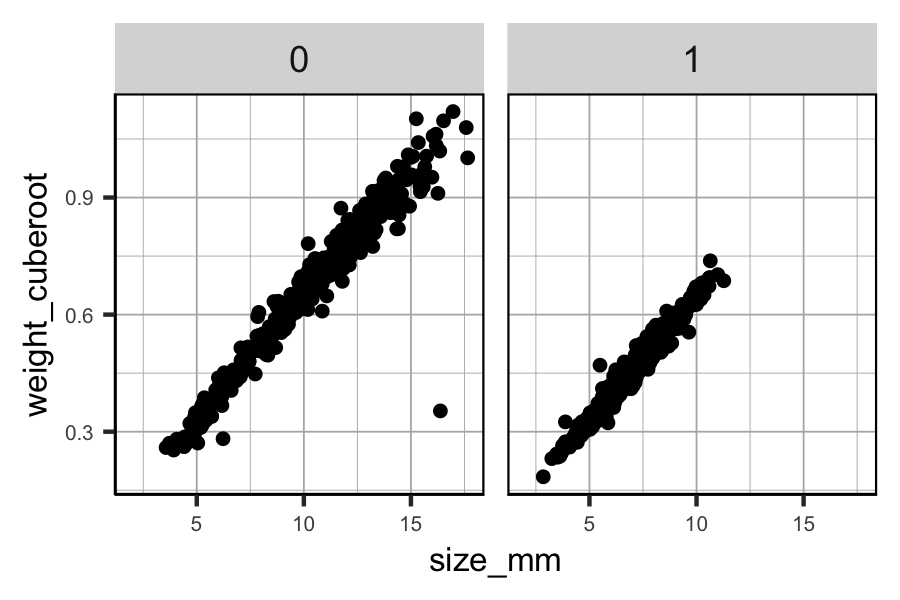


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

##   
## [[4]]

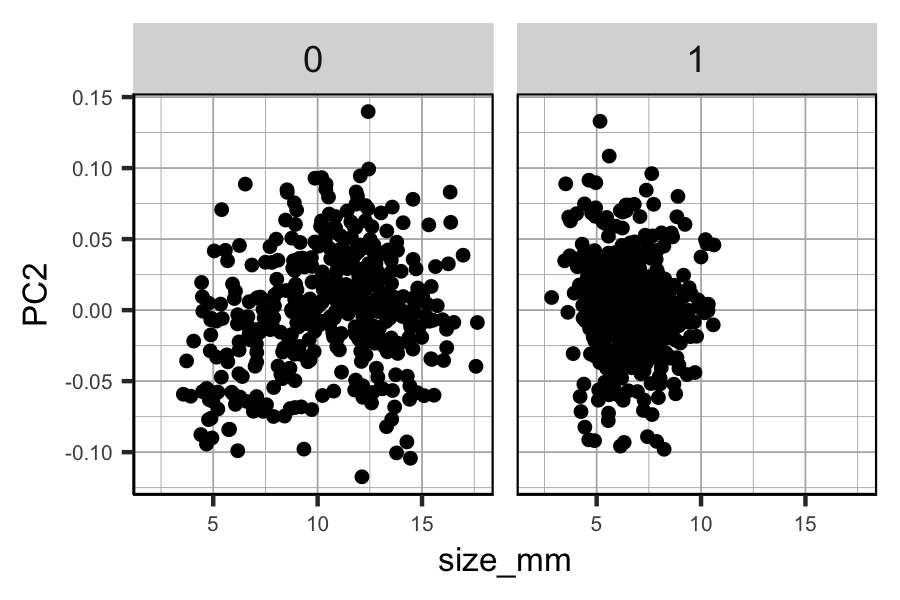


Figure 1. Scatter plots of the relationship between size and the other phenotypes in each generation separately.

None of the diagonal elements are close to zero but the closest values are found in size, thickness and weight. Given these results, we can exclude mutlicollinearity and focus on the consequences of using highly correlated variables in the model. However, I am quite skeptical about the abscence of multicollinearity because in general, it is expected that an absolute correlation coefficient of among two or more predictors indicates the presence of multicollinearity (Dohoo et al. 1997). Perhaps I am performing the Choleski decomposition of the correlation matrix on the wrong set of variables. What would you suggest?

Dohoo, I., Ducrot, C., Fourichon, C., Donald, A. and Hurnik, D. (1997), “An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies”, Preventive Veterinary Medicine, Vol. 29 No. 3, pp. 221-239.

## Part two: Size at maturity

Just briefly because I think we should clarify part one first but for this section, the plan is to perform an ordinal logistic regression between the multinomial outcome variable (mature, immature and juvenile) and the predictor variable size. The result should give us an estimate of the probability of maturation depending on size. Not sure about how to split the data though. By generation and by tanks or just by tanks? I guess that in the first case we might be able to see whether the probablity of maturation differ between generations and not only between tanks and thus ecotypes but there are compromises to make.

1. In the generation 0 parents there is not enough information about maturity (Table 2) but we could use CZA data from the contact zone experiment.
2. CZA data were collected continuously along the shore and not discretely by patches but we could imitate the sampling that was done for the offspring experiment and select CZA individuals within a circle with center equal to the mean path of each population and radius of 2m (or maybe a different radius?).

Does it make sense to try by “generation” and by tanks?

I agree about the experimenter problem here! It may also be a problem for males - some people are better at spotting a small penis than others. It would be nice to do separate analyses for males and females but we cannot distinguish sex of juveniles so that is not possible (at the moment - maybe markers near the sex determination locus will eventually make this possible…).

The first thing I would do would be to plot size x tank and colour by the classes: juvenile, imm.male, imm.female, male, female. Then we get a visual impression of size at maturity and how it changes over tanks.

Then we could think about fitting a model where the probability of maturity changes with size. If we ignore ‘immature’ individuals for the moment (probably treat them as juvenile), we can do something like this:

p\_m and p\_f are the probabilities that a male or a female will be adult, given their size, the average size at which males or females become adult (smat\_m, smat\_f) and a slope (b\_m, b\_f) that determines how variable the size at maturity is.

Then we also have the unknown sex ratio (sr) in juveniles so a snail is juvenile with probability

maybe we don’t need this because it should just be 1-p\_m-p\_f (any snail that is not male or female is juvenile)

That means we have 4 unknowns that we can fit given the size and maturity status of the individuals and assuming the parameters are constant across the contact zone. After fitting this model, it could be extended to allow the parameters to vary over tanks (or with space in the clinal data). I did something a bit like this ages ago for ANG, but without separating males and females. It was possible to fit a cline for size at maturity even though we had rather few juveniles in that data set.