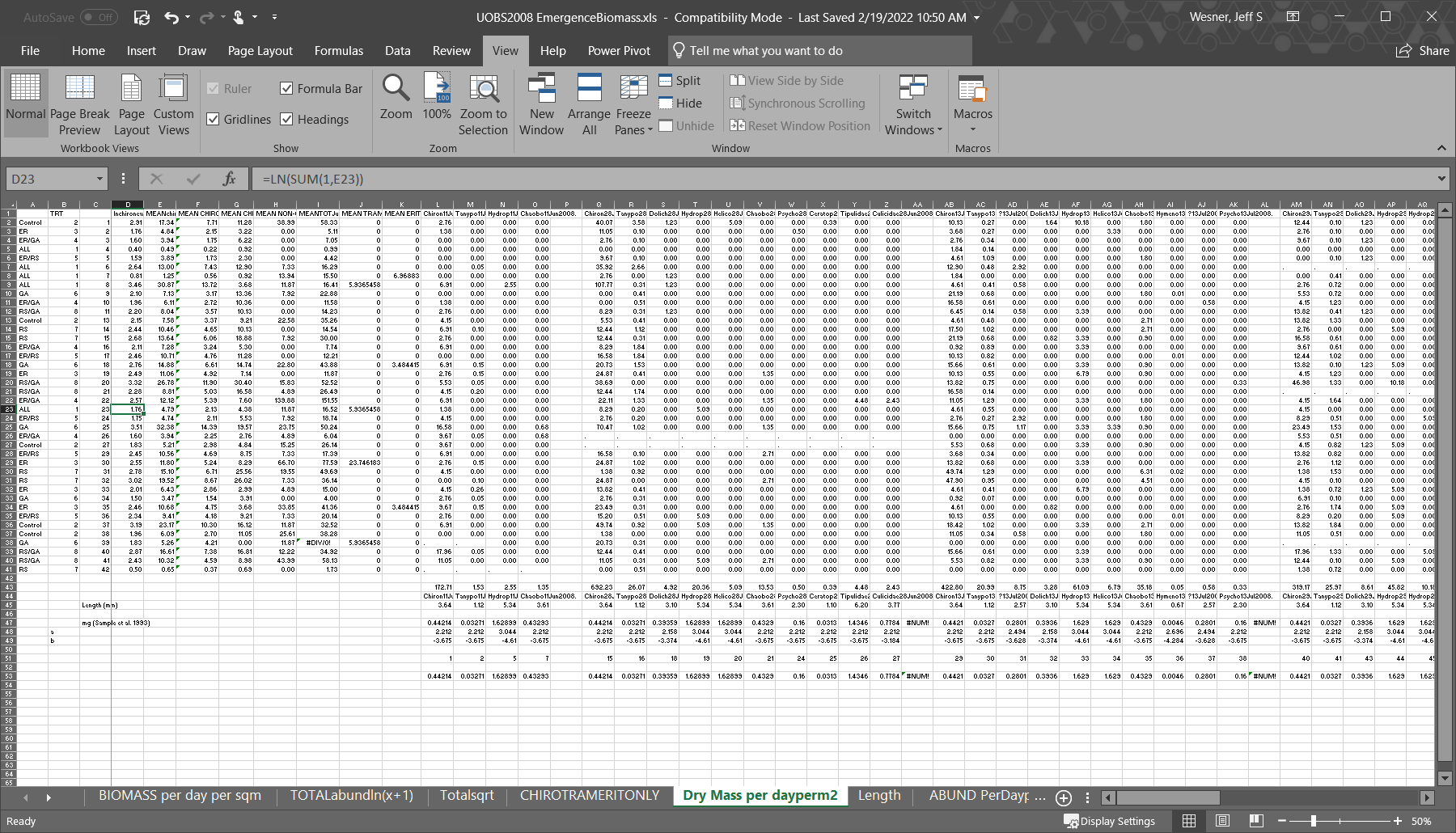
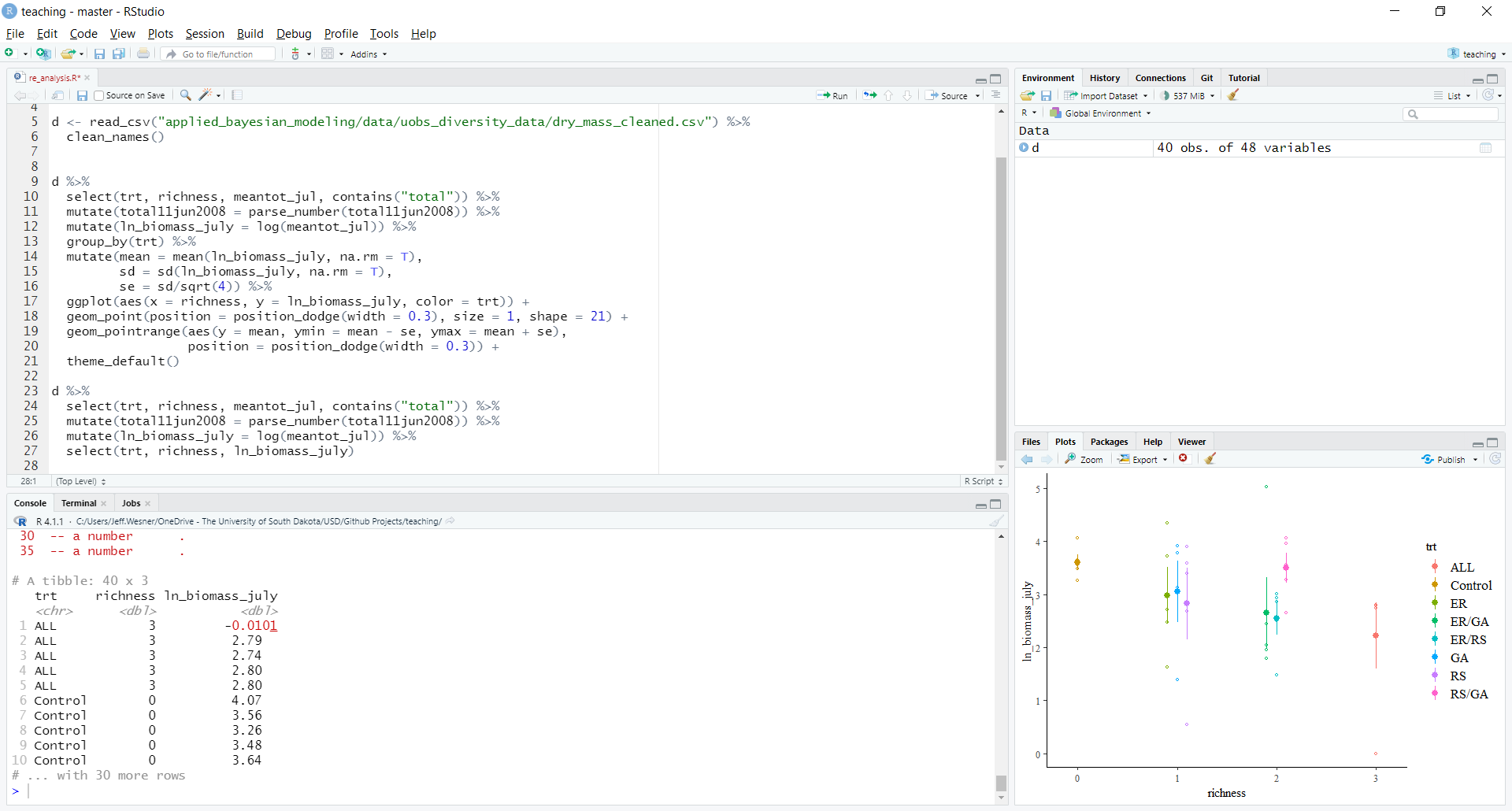
Recreate Figure 1a from Wesner 2012 (Oikos).

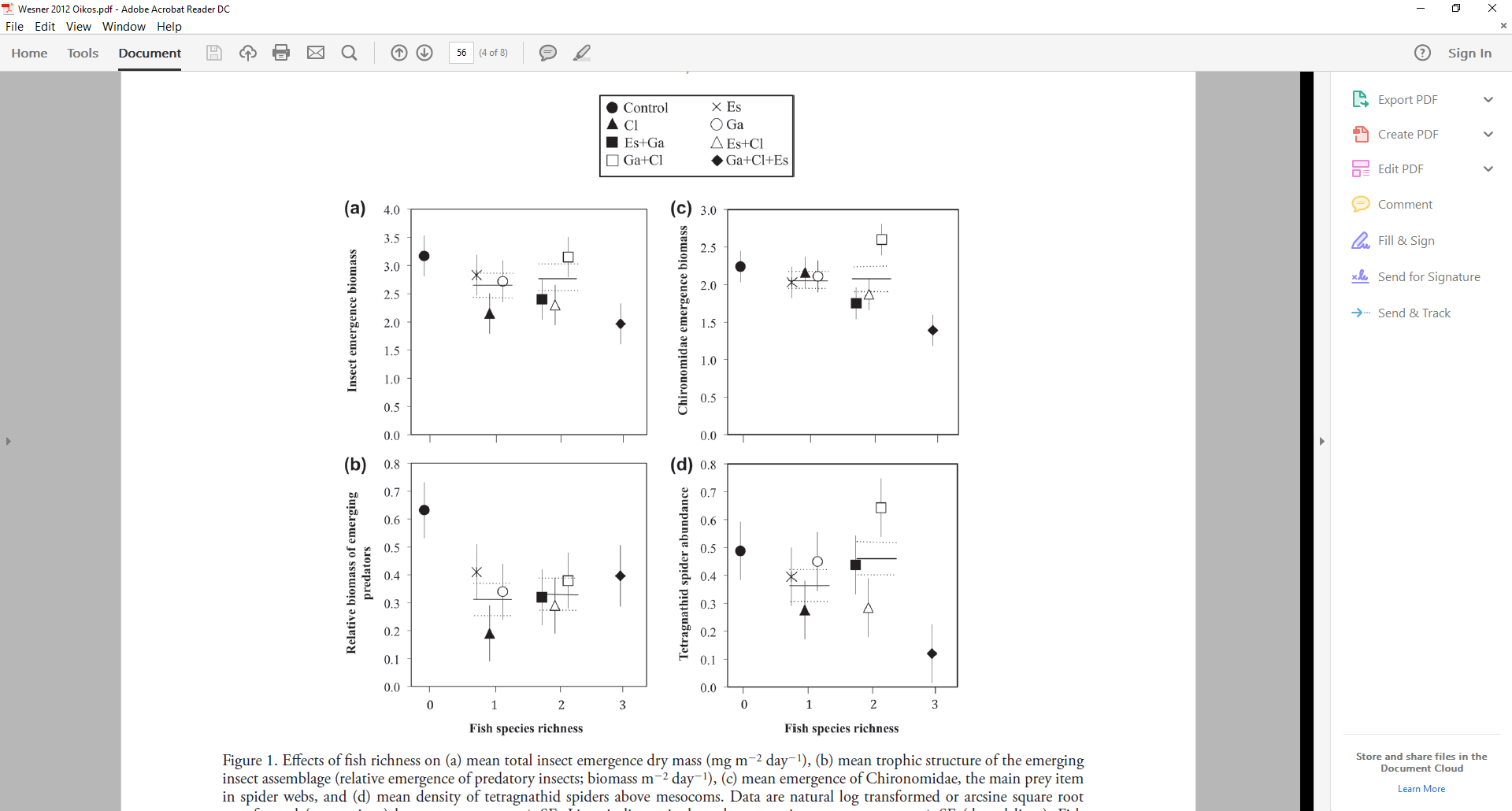
Step 1: Can I figure out the data?

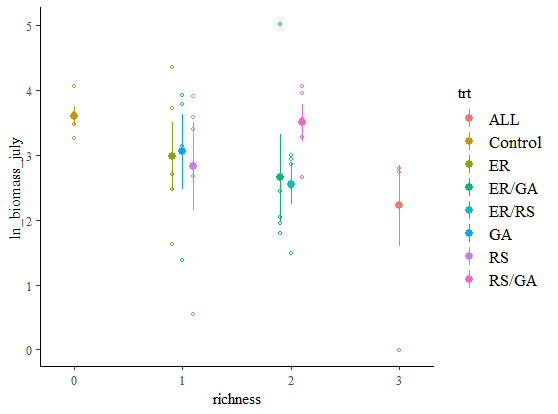


The original data are a bit of a mess. The whole analysis requires 3 columns of data (treatment, richness, and biomass). In the original data, I had 121 columns! Those columns represented every prey taxon, so that makes some sense. But they also represent a ton of exploratory data, like log transforming everything and arcsin square root transforming proportions, etc. The dataset also contains formulae and length-weight regression at the bottom. This would be an awful dataset to send to someone. But the real data *are* in there. They look like this:



Step 2: Can I make this plot again?





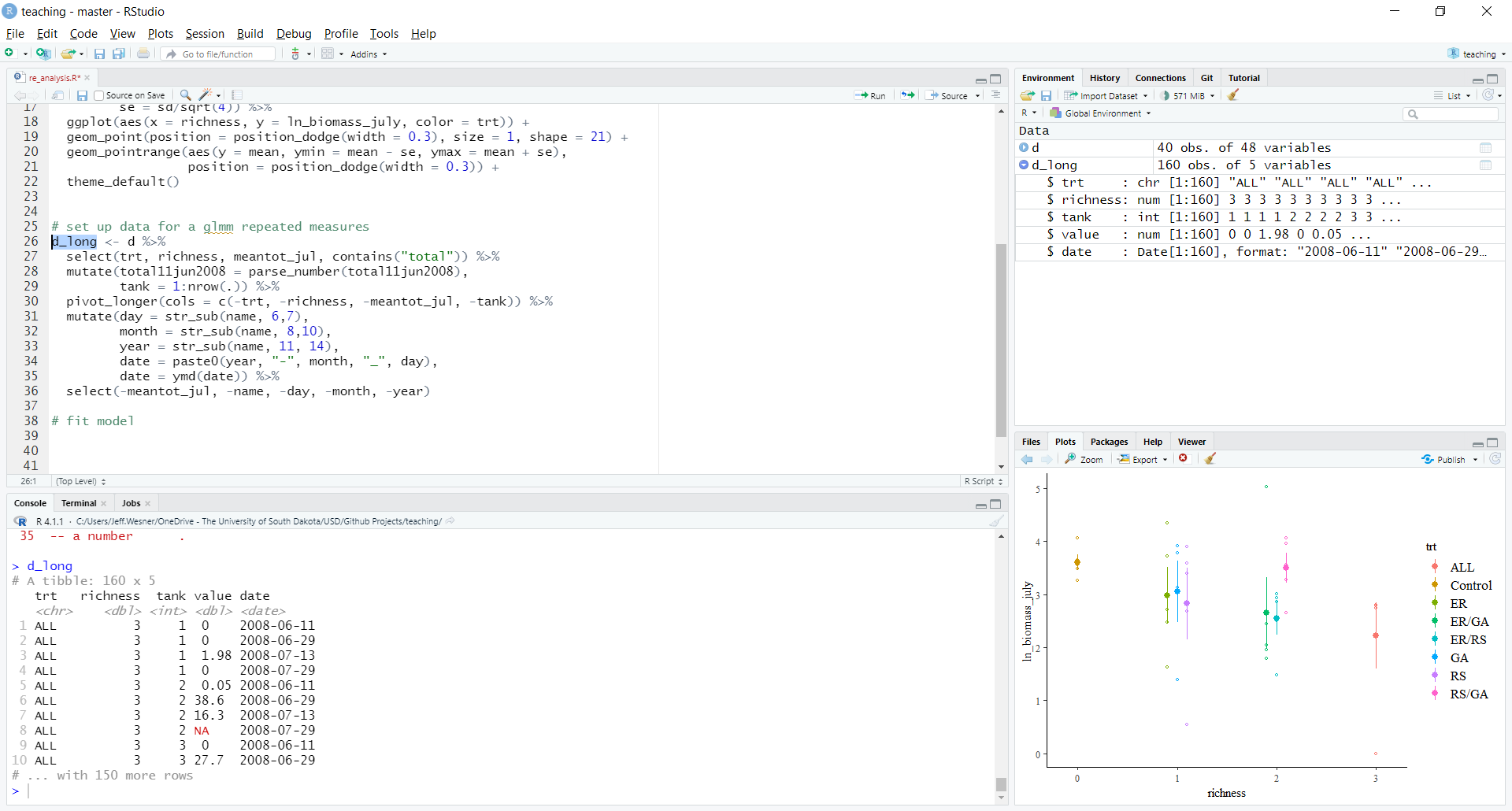
The data look about right. In the original paper, the raw data were not plotted. In the new plot, the raw data are shown. The means are exact, because the original data had least squares marginal means. I’m not sure how to calculate those now, so the plotted means are just raw arithmetic means of the data (not modeled means. But it all looks close enough. I’ll continue with this data set.

Another difference is the shapes. Ggplot gives a warning when there are more than 6 shapes, so I couldn’t recreate it exactly with ggplot.

Step 3: Is there enough information to re-fit the model? Here’s the original. Note that the analysis included time, even though that isn’t in the plots. I’m not sure how the plots and lsmeans came from the time analysis. But the data for time are available, so we can re-fit this as described, but using Bayes. The model also mentions that it is mixed, but doesn’t say what the random effect was. I’ll assume it was the mesocosm as a varying intercept, since that was the level of repeated measures (i.e., repeated measures from the mesocosms). There is no mention of the likelihood, but given the tests for normality, we can assume that it was Gaussian.



So the actual data to analyze (with time) look like this. The brms code is below that:



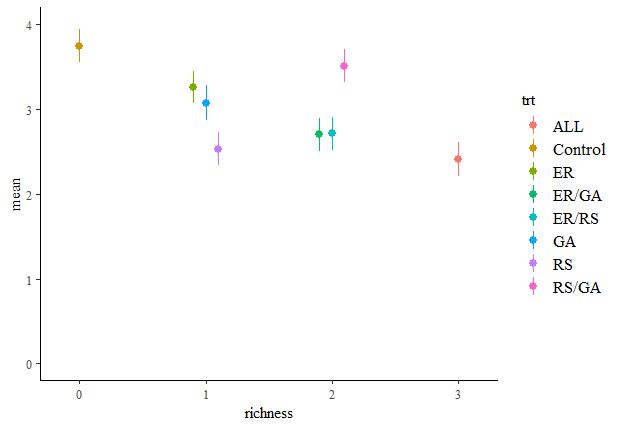
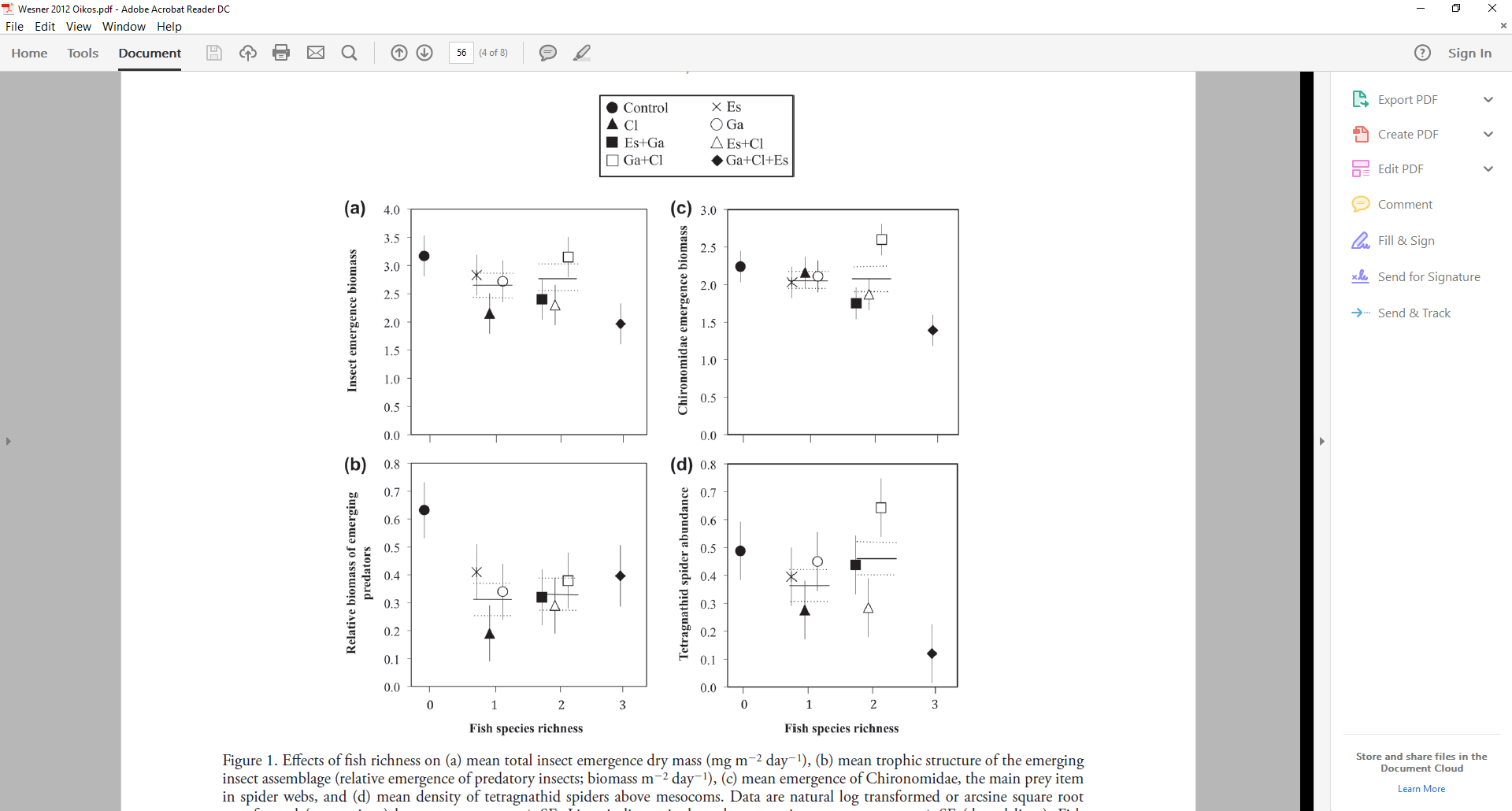
*brm\_uobs <- brm(value ~ trt\*date + (1|tank),*

*family = gaussian(),*

*data = d\_long)*

Here’s what I would write today

*To analyze the effects of fish treatments on insect emergence, I fit a general linear mixed model with natural-log transformed insect emergence as the response variable, time, treatment and their interaction as predictor variables, and mesocosm number as a random intercept. The likelihood was gaussian. To compare differences among treatments, I used derived quantities from the posterior distribution of this model to estimate treatment differences among treatment combinations. These were compared by first averaging emergence in each treatment over the last three collection dates and then estimating differences across each iteration of the posterior distribution.*



On the left is the model results from the brms model. They look pretty close to the original. Phew! Here’s what I wrote about them:



Here’s what I would write today.  
*Mean log-emergence dry mass was 23% (95% CrI: 2 to 37%) lower in pools with fish than the control treatments, with a 98% probability that the difference was greater than zero. The three species treatment caused a 36% (9 to 58%) reduction in log-emergence compared the control, with a 99% probability that the difference was greater than zero. Pools with one or two fish species reduced insect emergence 11% and 10%, respectively, with a >95% probability that the difference was greater than zero. Across all fish treatments, there was a >89% probability that the treatment with three species had lower emergence than the treatments with either one or two species, but there was also an >11% probability that it had higher emergence.*

Here’s what I would change with the model overall if I were to re-run it today:

1. Use a Gamma likelihood on the raw data, instead of log-transforming the data and using gaussian
   1. This does two things. First, it uses a likelihood that assumes that the variance increases with the mean, which is true of most datasets
   2. It allows us to make inferences about emergence in units of dry mass, rather than log-dry mass.
2. Reduce the scale of the data to make the model run smoother. i.e., model grams dry mass instead of mg dry mass
3. Another thing I would try is to model richness itself as the predictor rather than treatments, instead using treatments as a random intercept. That more directly tests the main hypothesis that fish richness affects emergence.

I quickly ran a model doing 1 and 2 above

