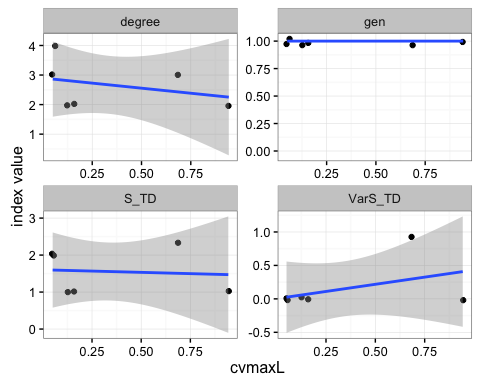
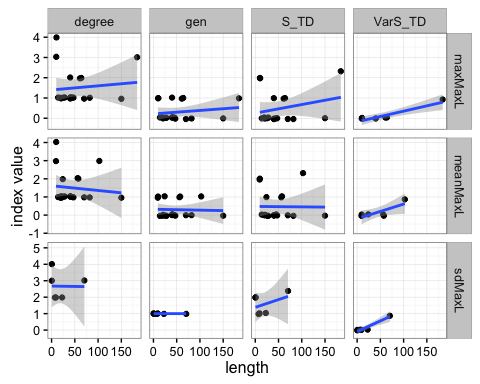
Generalism data results

Josephine Walker

14 May 2016

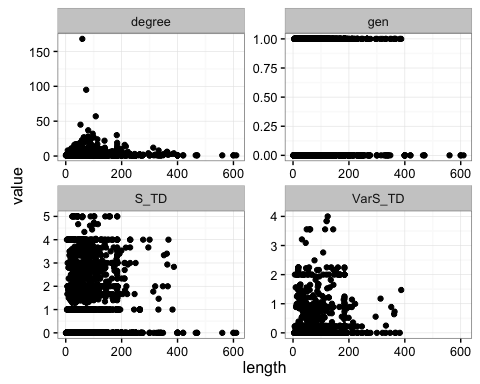
## Prediction 1: It is easier for generalist endoparasites to invade if the available hosts are large-bodied

I looked for a relationship between the mean, coefficient of variation () and maximum MaxL (fish host length) value for each parasite’s hosts and each of the four parasite generalism indices (*degree*, the number of hosts; *gen*, a binary of whether the parasite has more than one host; *STD*, the average taxonomic distance between the hosts; and *VarSTD*, the variance of *STD*).

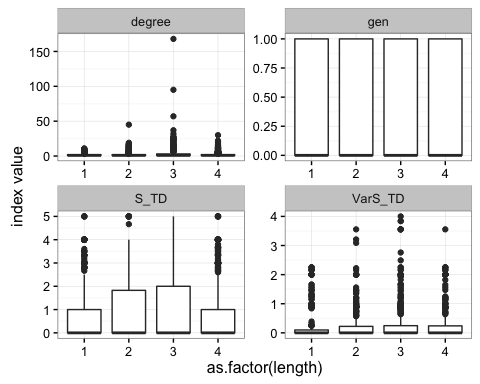


## Ectoparasites and host body size - more complex relationship expected

## Warning: Removed 2687 rows containing missing values (geom\_point).

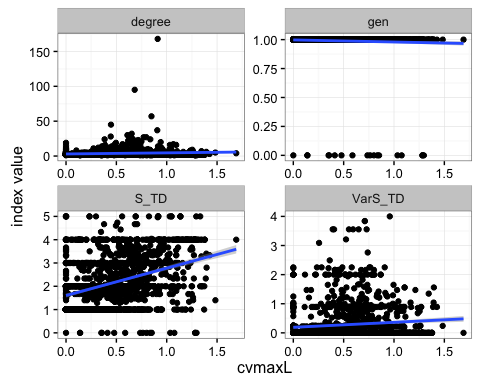


## Warning: Removed 2651 rows containing non-finite values (stat\_boxplot).



## Warning: Removed 10606 rows containing non-finite values (stat\_smooth).

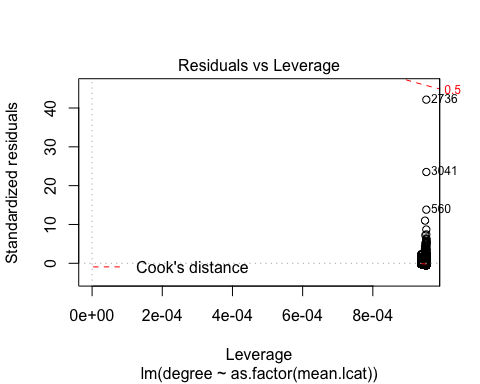
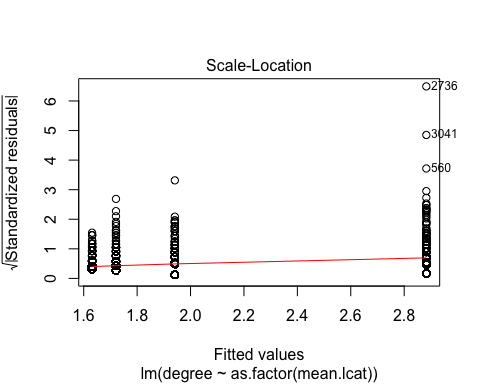
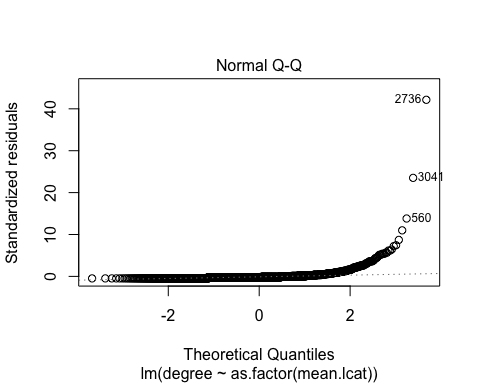
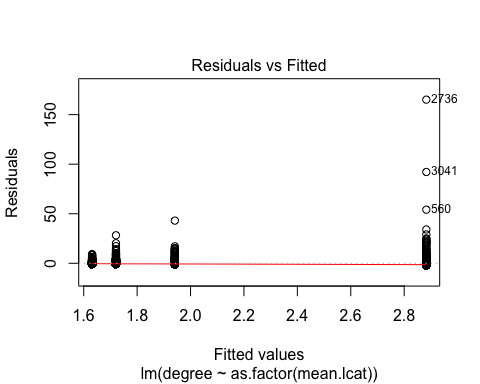
## Warning: Removed 10606 rows containing missing values (geom\_point).



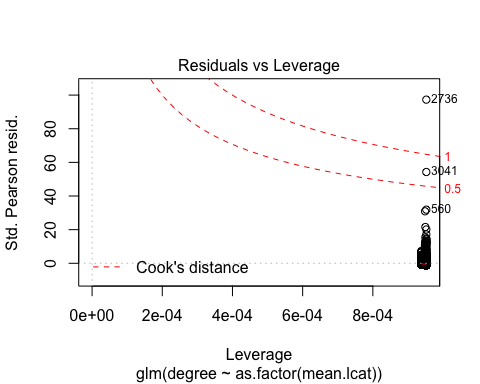
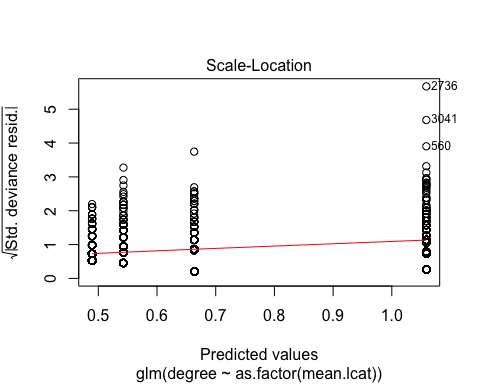
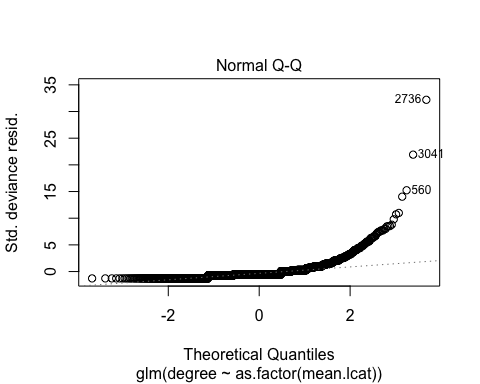
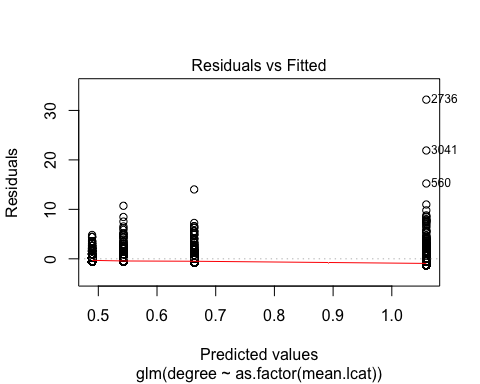
## Source: local data frame [16 x 4]  
## Groups: index [?]  
##   
## index length mean median  
## (chr) (dbl) (dbl) (dbl)  
## 1 degree 1 1.6313321 1  
## 2 degree 2 1.9411765 1  
## 3 degree 3 2.8828571 1  
## 4 degree 4 1.7207977 1  
## 5 gen 1 0.3386492 0  
## 6 gen 2 0.3814042 0  
## 7 gen 3 0.4800000 0  
## 8 gen 4 0.2896486 0  
## 9 S\_TD 1 0.7724211 0  
## 10 S\_TD 2 0.8468111 0  
## 11 S\_TD 3 1.0655333 0  
## 12 S\_TD 4 0.5950081 0  
## 13 VarS\_TD 1 0.2338901 0  
## 14 VarS\_TD 2 0.2296202 0  
## 15 VarS\_TD 3 0.2982014 0  
## 16 VarS\_TD 4 0.3337894 0

##   
## Kruskal-Wallis rank sum test  
##   
## data: degree by as.factor(mean.lcat)  
## Kruskal-Wallis chi-squared = 100.84, df = 3, p-value < 2.2e-16

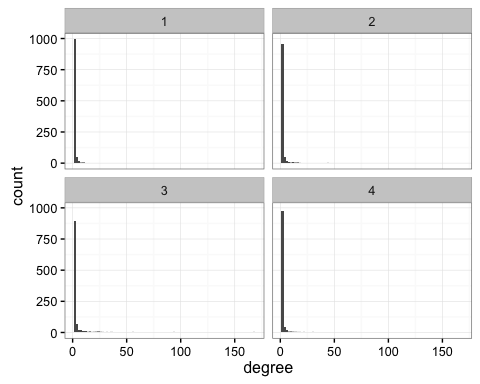
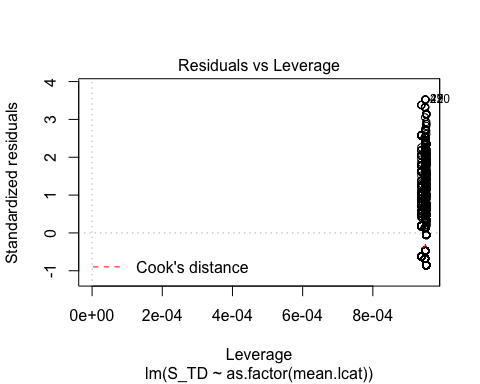
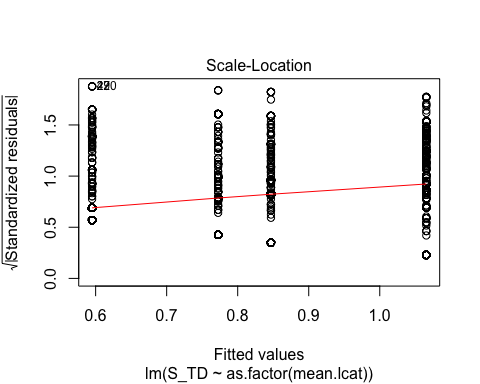
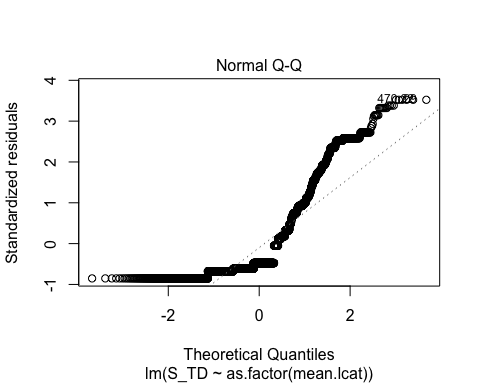
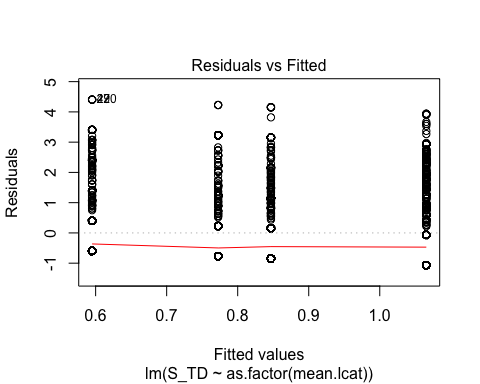
##   
## Call:  
## lm(formula = degree ~ as.factor(mean.lcat), data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.883 -0.941 -0.721 0.117 165.117   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.63133 0.12003 13.591 < 2e-16 \*\*\*  
## as.factor(mean.lcat)2 0.30984 0.17023 1.820 0.0688 .   
## as.factor(mean.lcat)3 1.25153 0.17039 7.345 2.46e-13 \*\*\*  
## as.factor(mean.lcat)4 0.08947 0.17027 0.525 0.5993   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.919 on 4219 degrees of freedom  
## Multiple R-squared: 0.01582, Adjusted R-squared: 0.01512   
## F-statistic: 22.61 on 3 and 4219 DF, p-value: 1.625e-14



##   
## Call:  
## glm(formula = degree ~ as.factor(mean.lcat), family = "poisson",   
## data = hp.ectoDirect)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.284 -0.745 -0.551 0.069 32.182   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.48940 0.02398 20.408 < 2e-16 \*\*\*  
## as.factor(mean.lcat)2 0.17390 0.03262 5.332 9.73e-08 \*\*\*  
## as.factor(mean.lcat)3 0.56938 0.03009 18.923 < 2e-16 \*\*\*  
## as.factor(mean.lcat)4 0.05339 0.03357 1.590 0.112   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 8350.2 on 4222 degrees of freedom  
## Residual deviance: 7872.1 on 4219 degrees of freedom  
## AIC: 17870  
##   
## Number of Fisher Scoring iterations: 6



##   
## Call:  
## lm(formula = S\_TD ~ as.factor(mean.lcat), data = hp.ectoDirect)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.0655 -0.8468 -0.5950 0.6011 4.4050   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.77242 0.03832 20.157 < 2e-16 \*\*\*  
## as.factor(mean.lcat)2 0.07439 0.05435 1.369 0.17113   
## as.factor(mean.lcat)3 0.29311 0.05440 5.388 7.5e-08 \*\*\*  
## as.factor(mean.lcat)4 -0.17741 0.05436 -3.264 0.00111 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.251 on 4219 degrees of freedom  
## Multiple R-squared: 0.01781, Adjusted R-squared: 0.01711   
## F-statistic: 25.51 on 3 and 4219 DF, p-value: 2.393e-16



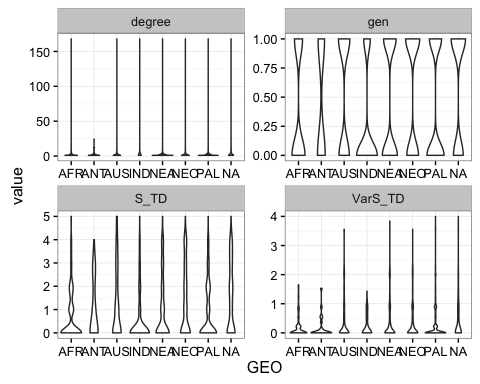
##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

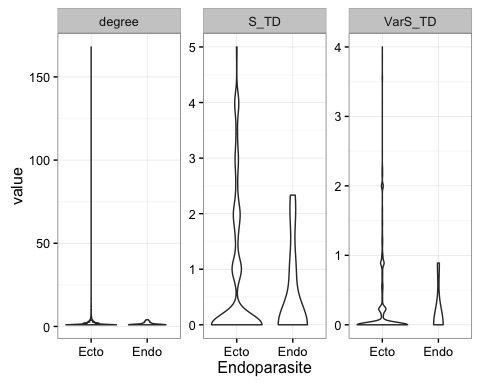
Negative binomial GLM seems like the best way to compare Ecto degree to quartiles of length categories. The estimates are log values, so exp() to get the estimate. Except, it's strictly positive so should use a zero truncated model?

Index isn't calculated based on geography, is it?

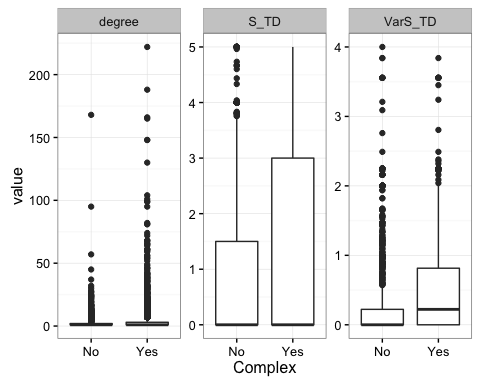
## Warning: Removed 2972 rows containing non-finite values (stat\_ydensity).



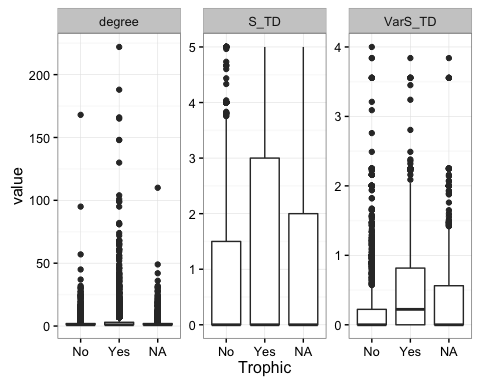
## Warning: Removed 2665 rows containing non-finite values (stat\_ydensity).



## Warning: Removed 4753 rows containing non-finite values (stat\_boxplot).



## Warning: Removed 6811 rows containing non-finite values (stat\_boxplot).



Maybe all the missing data is biased because it's for parasite species that are less well studied...