Branch length distributions.

Working notes.

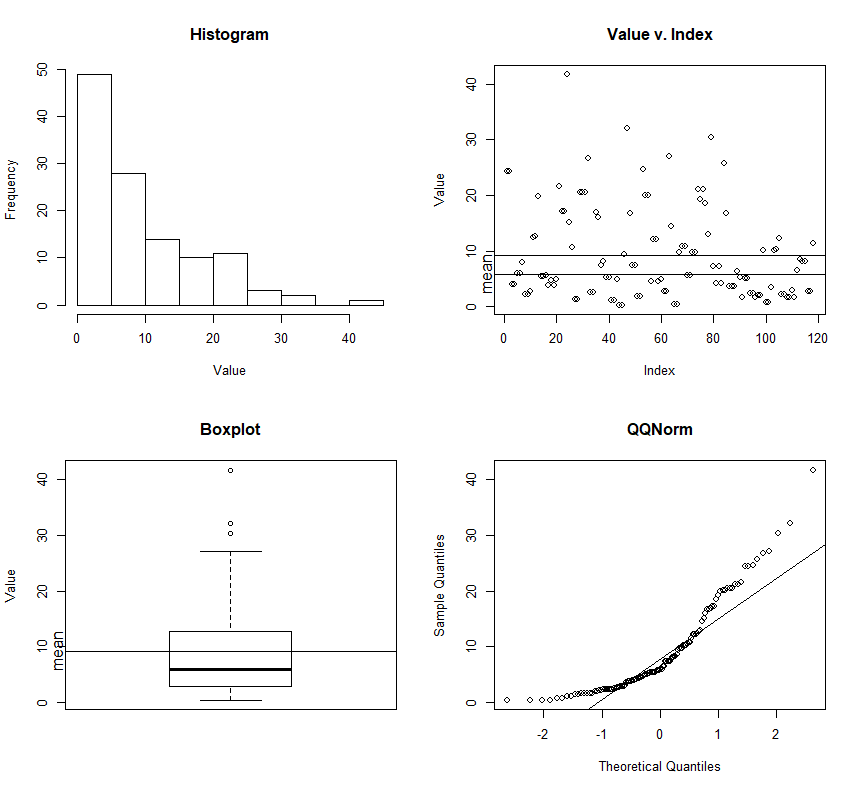
Leading number is number of taxa included.

118 GALLIFORMES

> summary(rttree[rttree$ORDER=="GALLIFORMES",4]);eda.wp((rttree[rttree$ORDER=="GALLIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.3171 2.7860 5.8180 9.1630 12.5900 41.7600

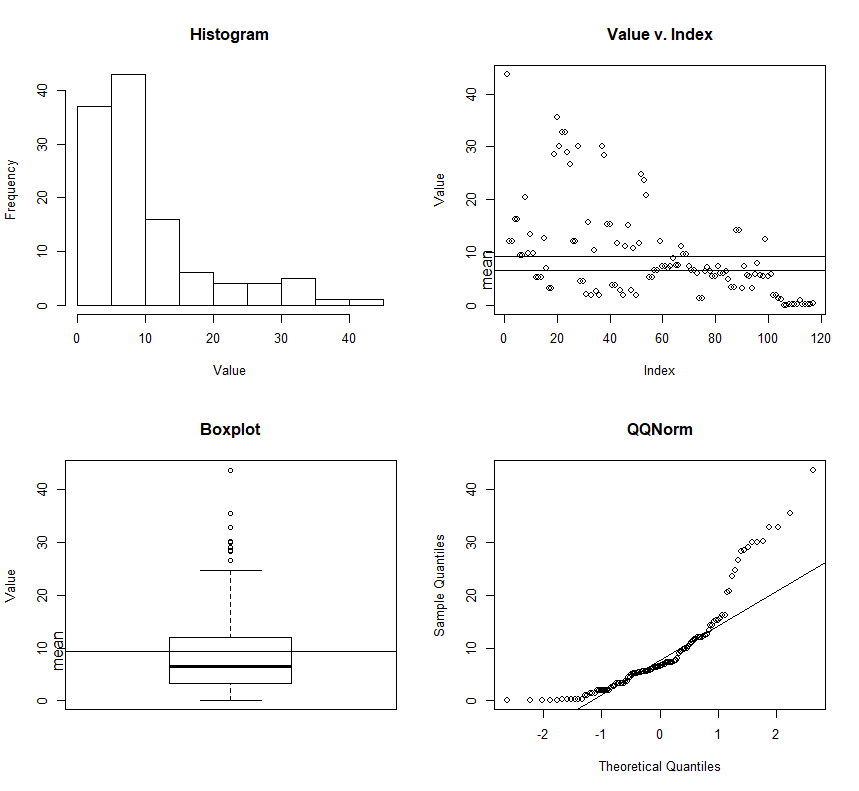


117 PASSERIFORMES\_oscines

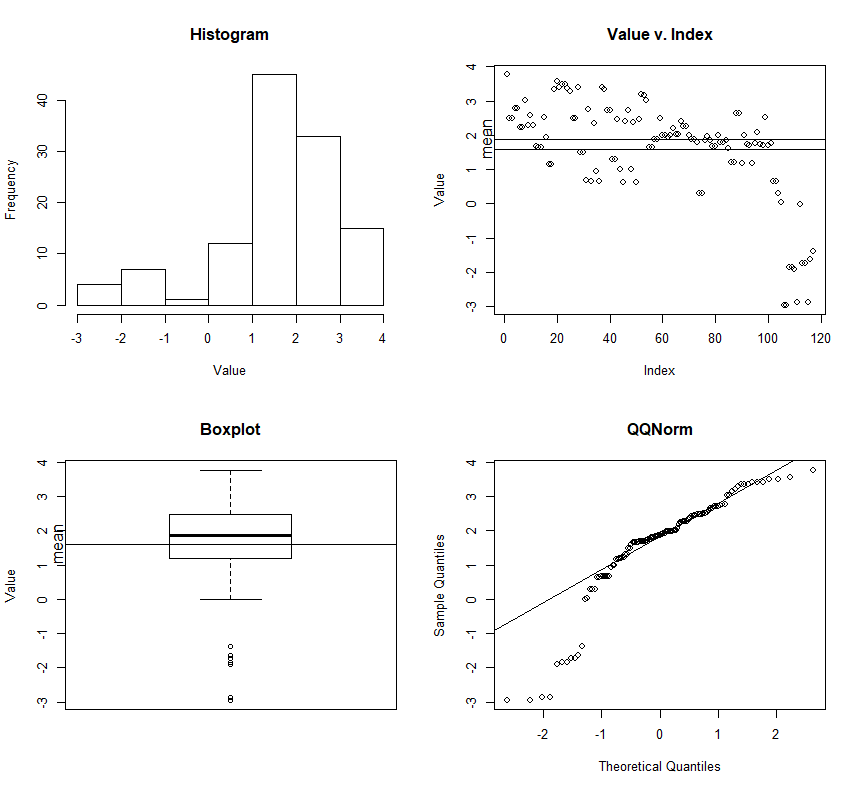
> summary(rttree[rttree$ORDER=="PASSERIFORMES\_oscines",4]);eda.wp((rttree[rttree$ORDER=="PASSERIFORMES\_oscines",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.05223 3.25800 6.50700 9.30700 12.08000 43.72000



LOG transformed.

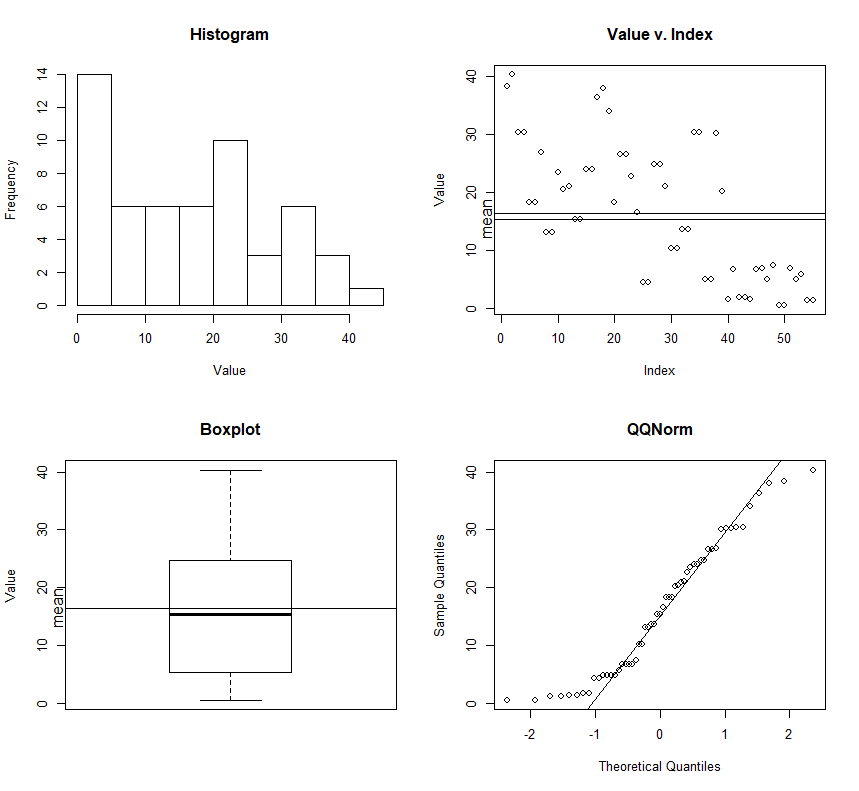


55 PASSERIFORMES\_suboscines

> summary(rttree[rttree$ORDER=="PASSERIFORMES\_suboscines",4]);eda.wp((rttree[rttree$ORDER=="PASSERIFORMES\_suboscines",4]))

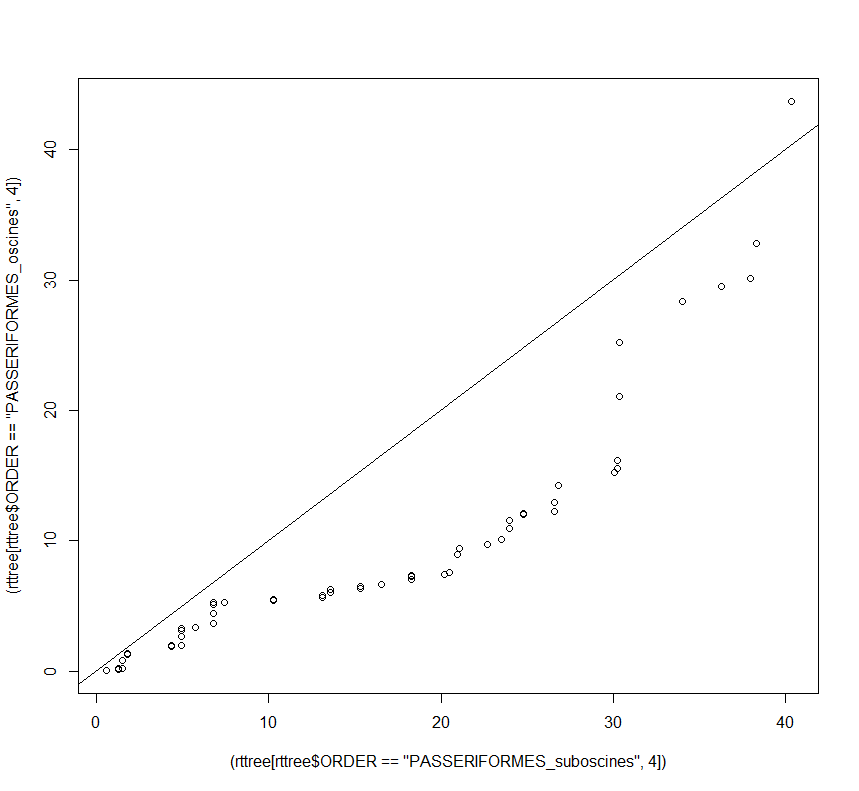
Min. 1st Qu. Median Mean 3rd Qu. Max.

0.5367 5.3250 15.3200 16.3600 24.7800 40.3500

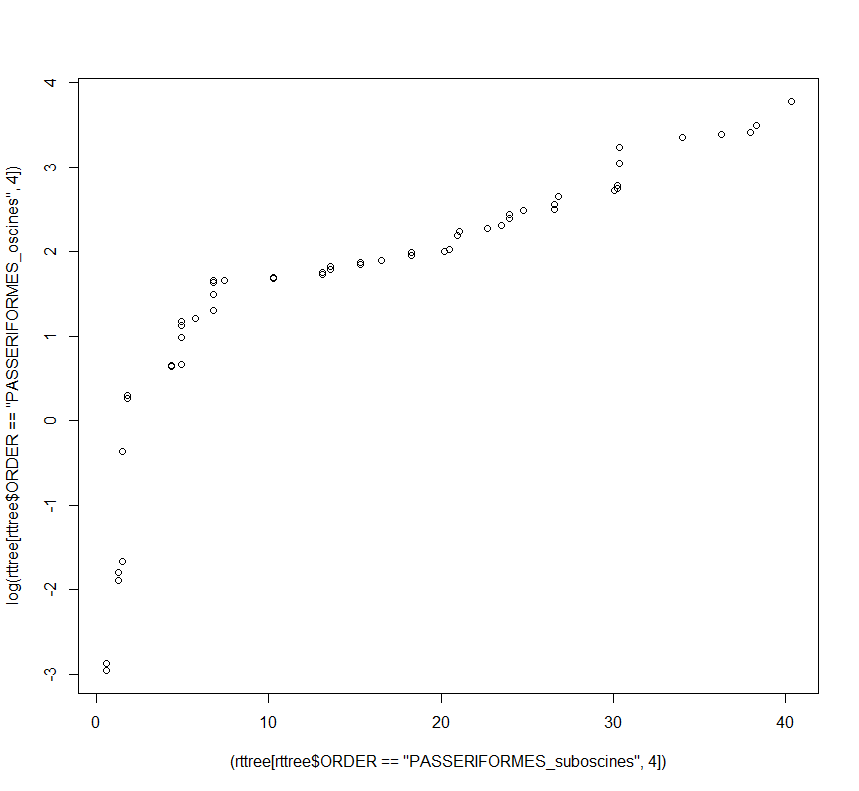


[APP: pretty well behaved without transformation… interesting]

Here is a qqplot doing a direct comparison of the suboscine and oscine subsets



THIS is weird, and I have no idea what (if anything it means, but the majority of the data becomes somewhat more “well behaved” if the suboscine data is log transformed. – this makes no sense to me, and I would ignore it (I think).



> summary(rttree[rttree$ORDER=="PSITTACIFORMES",4])

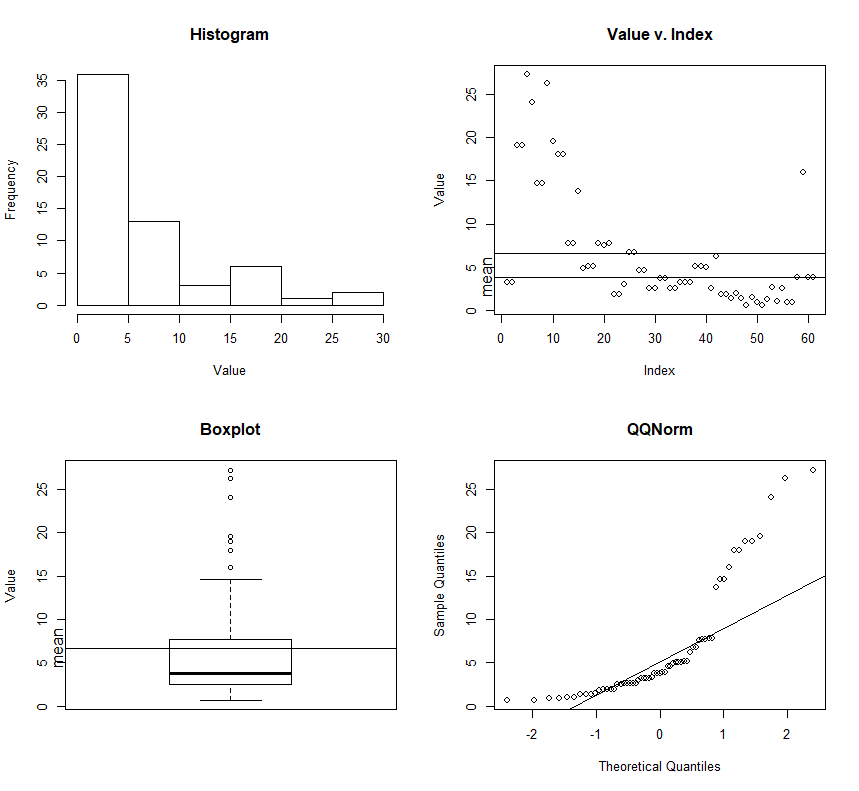
Min. 1st Qu. Median Mean 3rd Qu. Max.

0.6687 2.5410 3.8400 6.6420 7.7440 27.2600

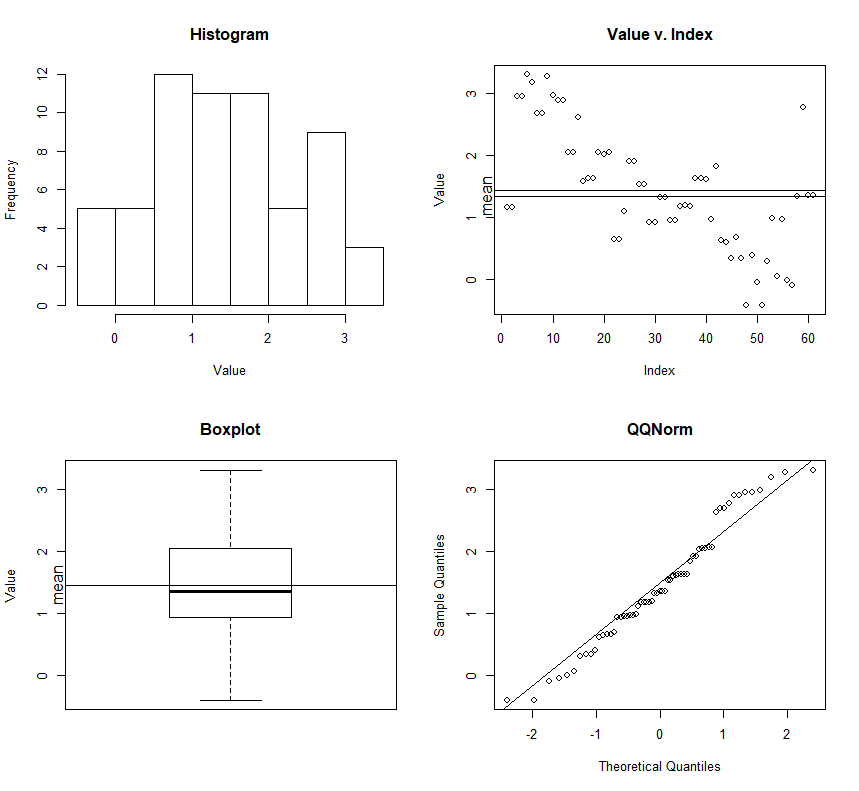
> summary(rttree[rttree$ORDER=="PSITTACIFORMES",4]);eda.wp((rttree[rttree$ORDER=="PSITTACIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.6687 2.5410 3.8400 6.6420 7.7440 27.2600



Implies log transformation would make sense.. here it is.



33 CORACIIFORMES

> summary(rttree[rttree$ORDER=="CORACIIFORMES",4])

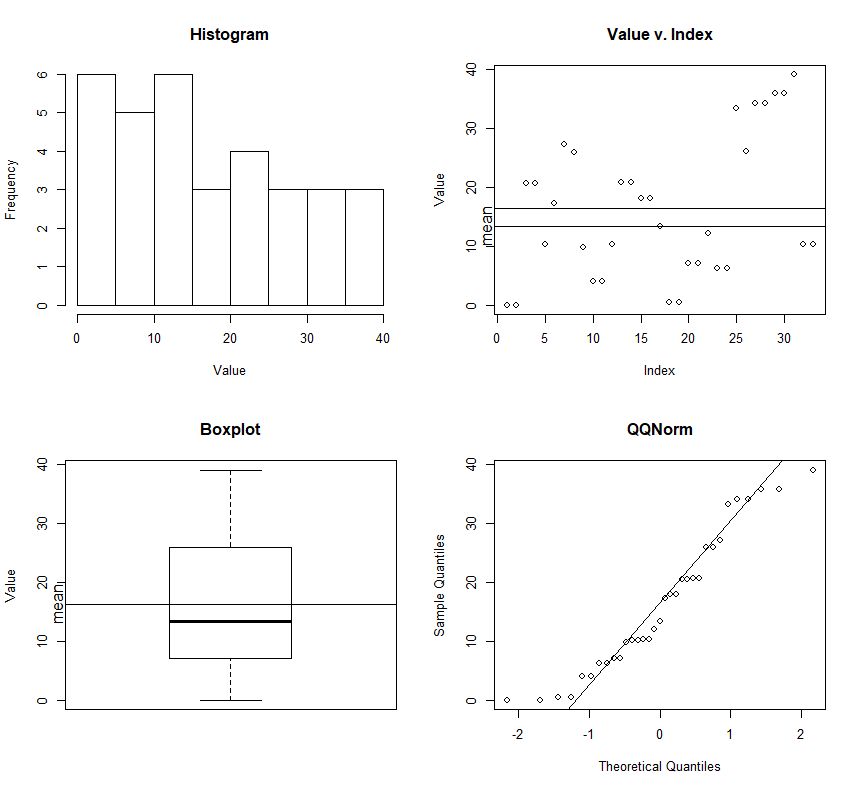
Min. 1st Qu. Median Mean 3rd Qu. Max.

0.04529 7.13900 13.39000 16.36000 25.91000 39.06000

> summary(rttree[rttree$ORDER=="CORACIIFORMES",4]);eda.wp((rttree[rttree$ORDER=="CORACIIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.04529 7.13900 13.39000 16.36000 25.91000 39.06000



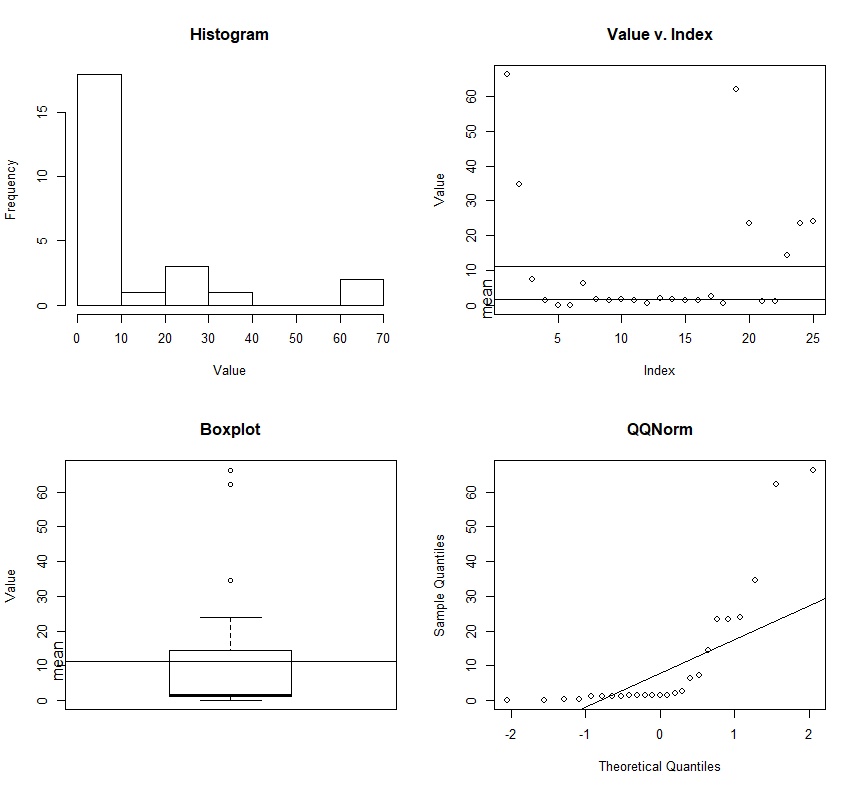
Pretty well behaved without log transformation.

25 ANSERIFORMES

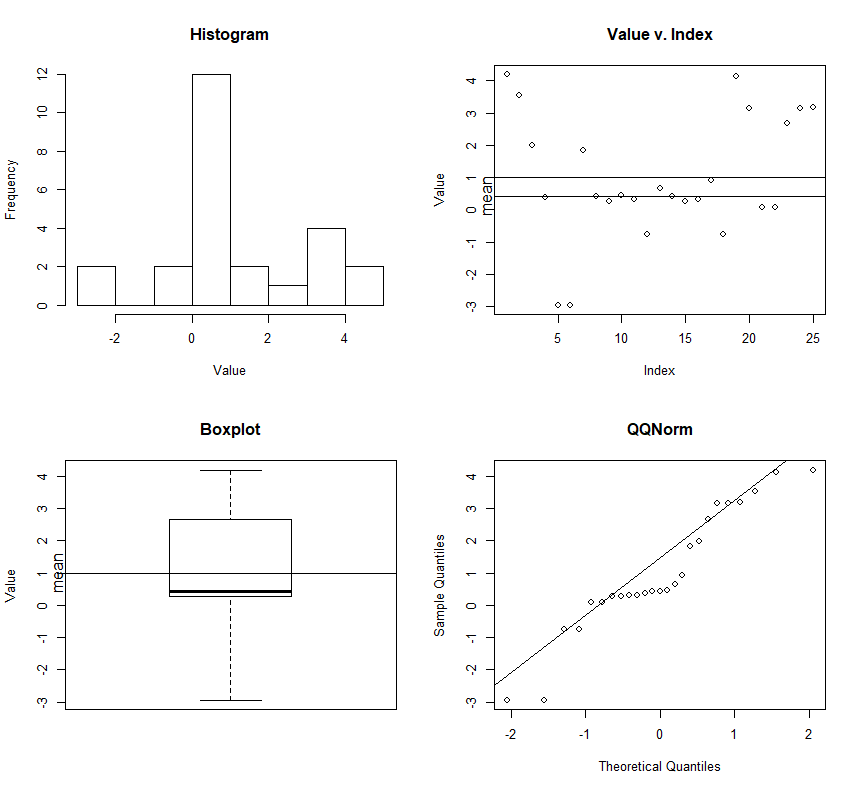
> summary(rttree[rttree$ORDER=="ANSERIFORMES",4])

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.05253 1.31100 1.53600 11.25000 14.39000 66.38000



Now log transformed.



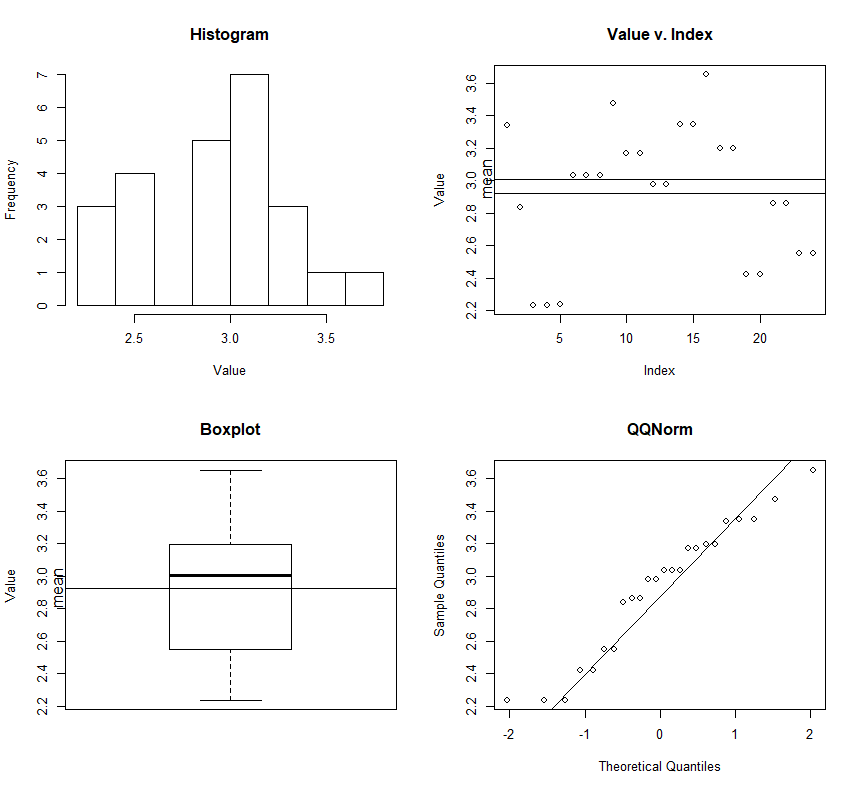
24 PICIFORMES

ry(rttree[rttree$ORDER=="PICIFORMES",4])

Min. 1st Qu. Median Mean 3rd Qu. Max.

9.336 12.790 20.180 20.070 24.410 38.560

Piciformes note log.

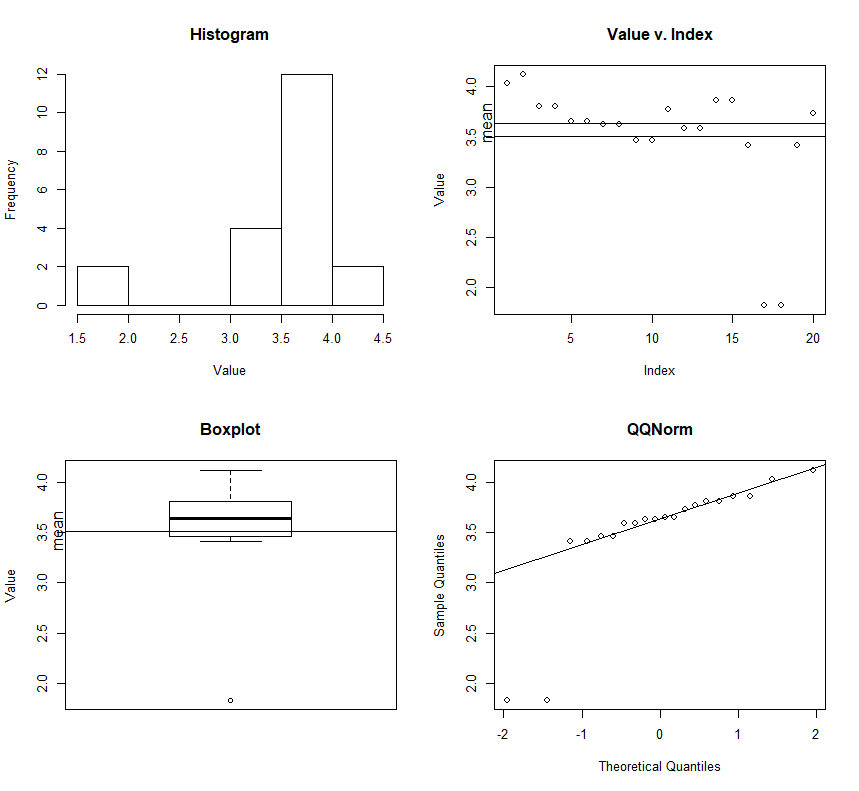


20 CHARADRIIFORMES

> summary(rttree[rttree$ORDER=="CHARADRIIFORMES",4])

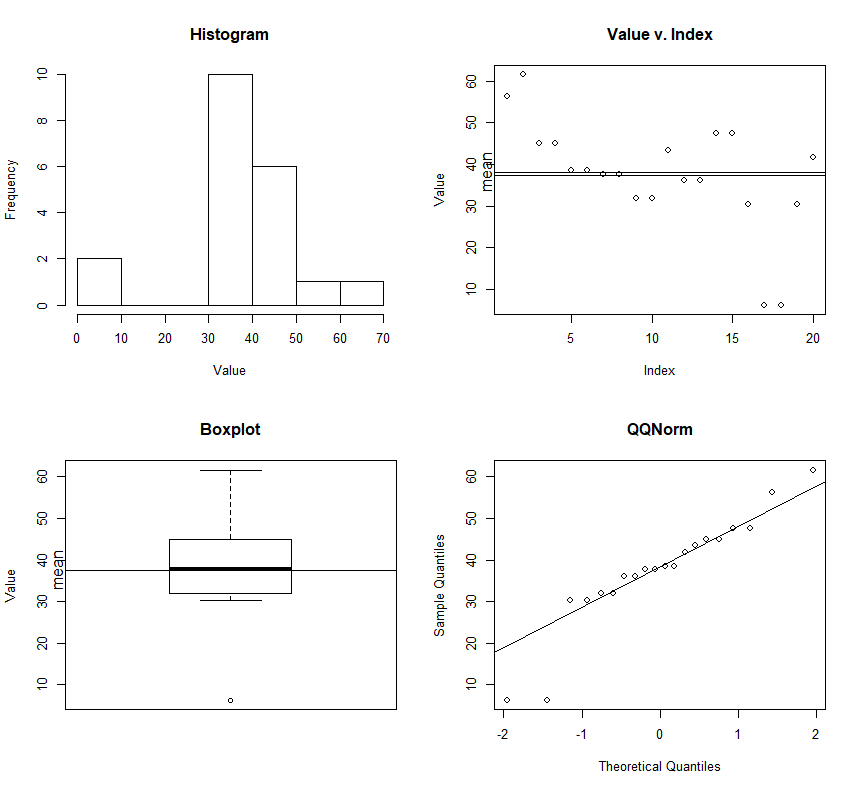
Min. 1st Qu. Median Mean 3rd Qu. Max.

6.235 31.840 38.060 37.450 44.940 61.620



[APP: well this looks WRONG!!!]

Here is non-log.



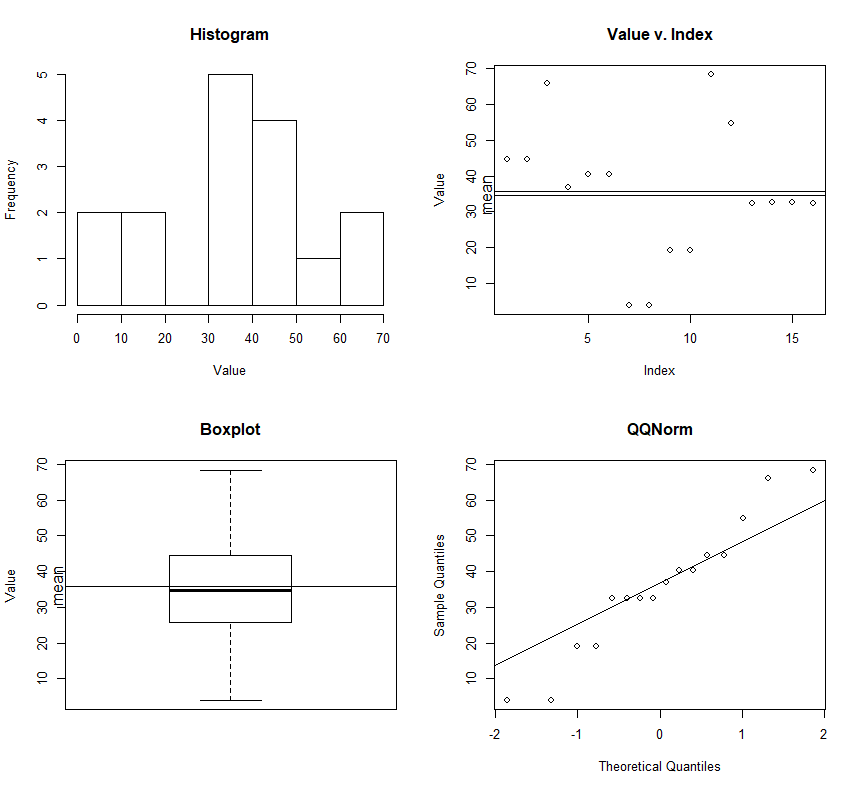
[APP: The low values certainly need investigation, but my goodness the branch lengths look long.]

16 CAPRIMULGIFORMES

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.905 29.130 34.700 35.750 44.600 68.440

> eda.wp((rttree[rttree$ORDER=="CAPRIMULGIFORMES",4]))



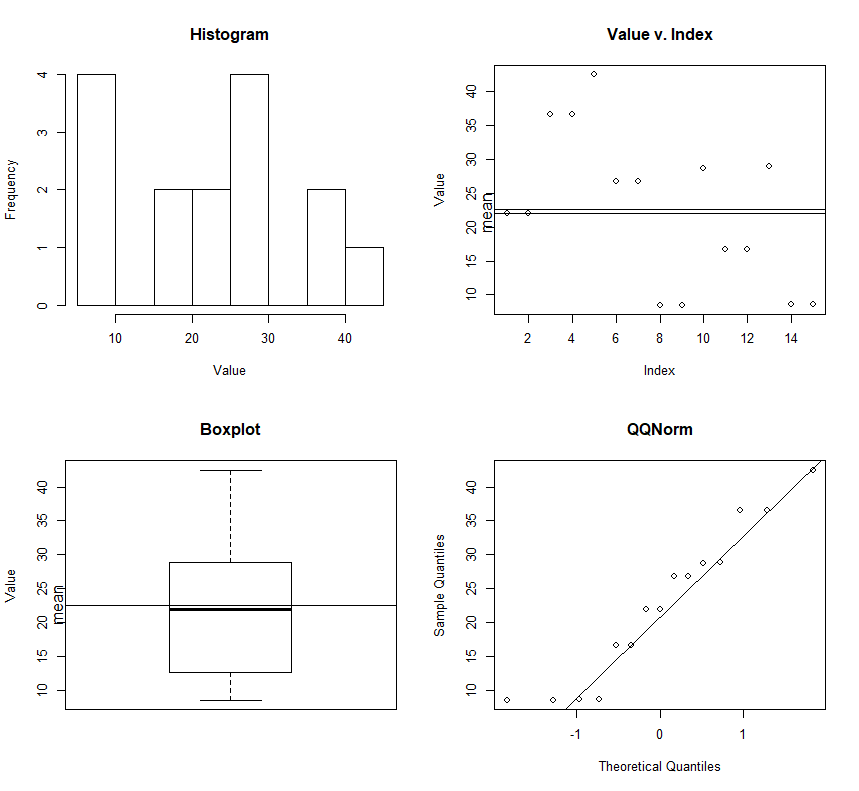
15 APODIFORMES

> summary(rttree[rttree$ORDER=="APODIFORMES",4])

Min. 1st Qu. Median Mean 3rd Qu. Max.

8.518 12.630 21.980 22.560 28.820 42.520

> eda.wp((rttree[rttree$ORDER=="APODIFORMES",4]))

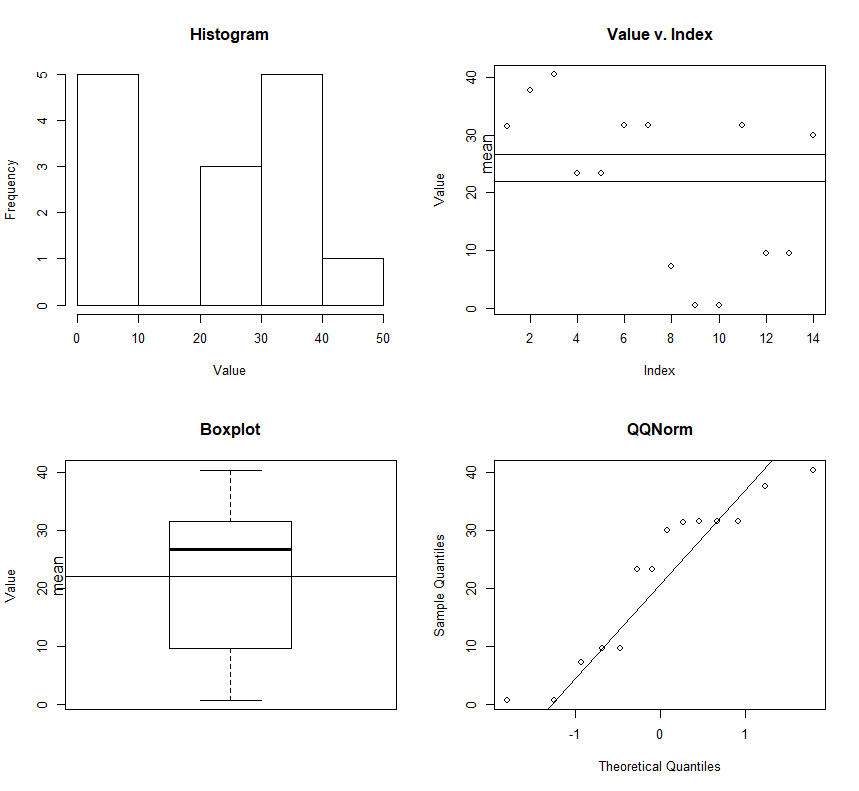


14 COLUMBIFORMES

> summary(rttree[rttree$ORDER=="COLUMBIFORMES",4]);eda.wp((rttree[rttree$ORDER=="COLUMBIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.6294 9.5750 26.6600 22.0600 31.5900 40.4800

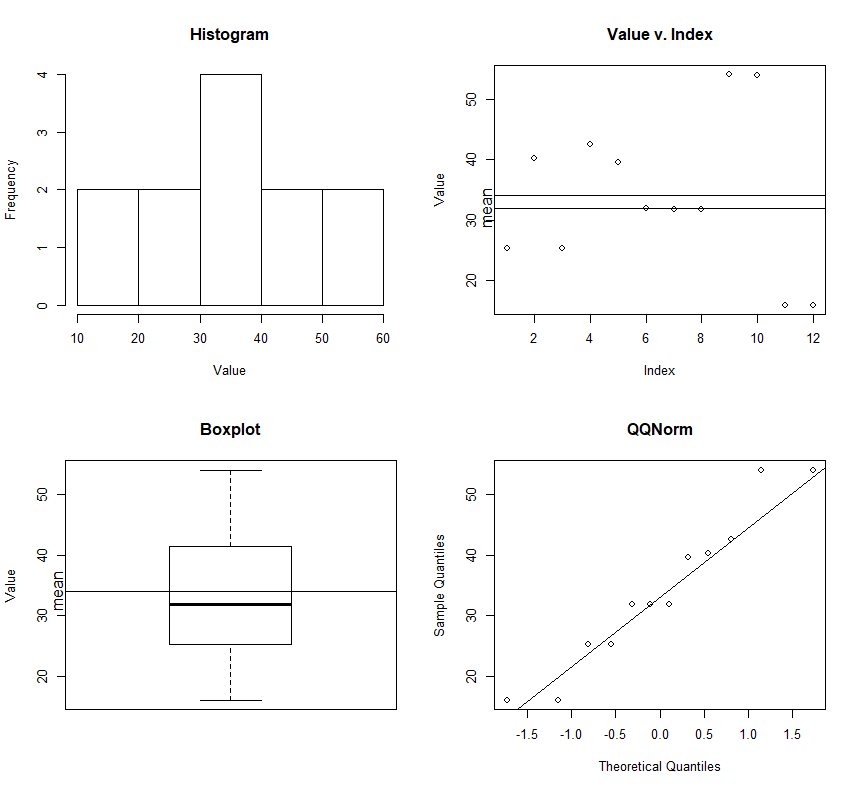


12 PELECANIFORMES

> summary(rttree[rttree$ORDER=="PELECANIFORMES",4]);eda.wp((rttree[rttree$ORDER=="PELECANIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

16.01 25.32 31.90 34.06 40.83 54.07

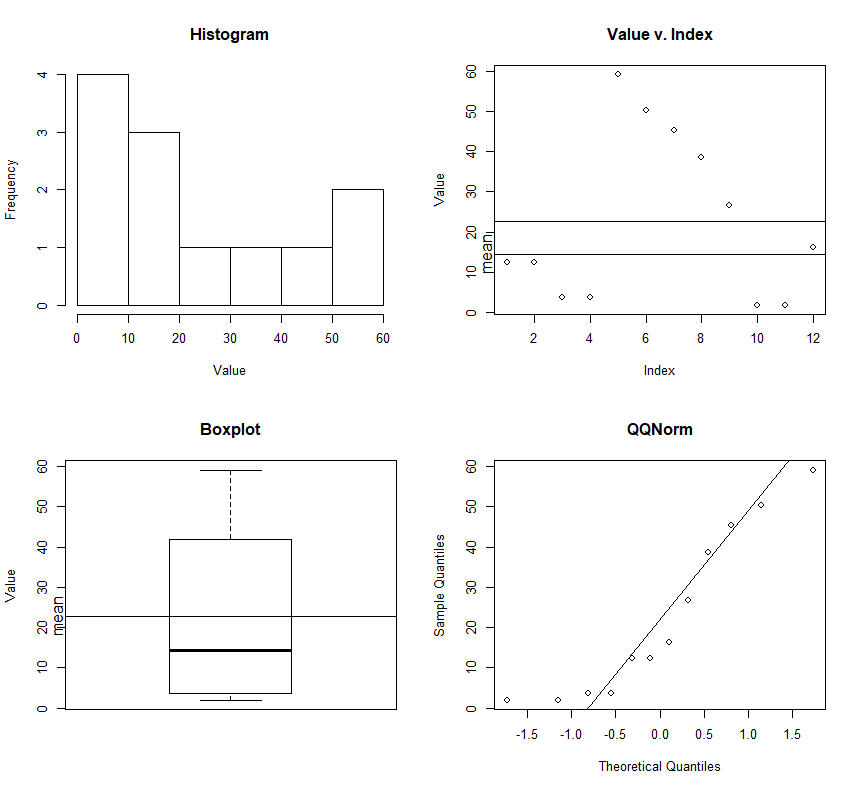


12 ACCIPITRIFORMES

> summary(rttree[rttree$ORDER=="ACCIPITRIFORMES",4]);eda.wp((rttree[rttree$ORDER=="ACCIPITRIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.882 3.642 14.340 22.680 40.270 59.180

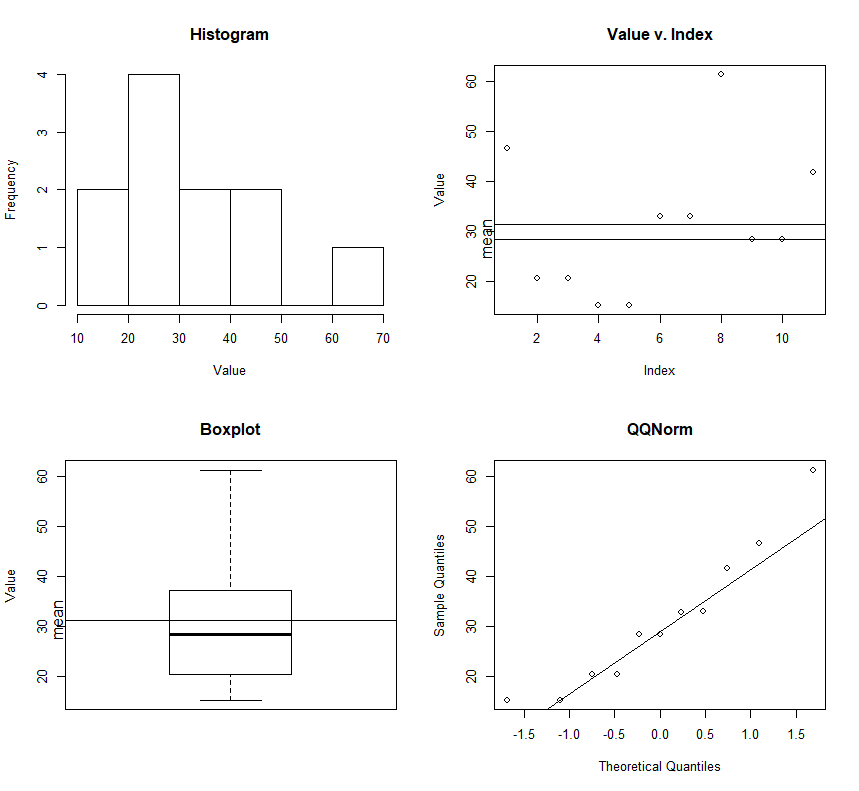


11 GRUIFORMES

> summary(rttree[rttree$ORDER=="GRUIFORMES",4]);eda.wp((rttree[rttree$ORDER=="GRUIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

15.25 20.54 28.38 31.25 37.31 61.29

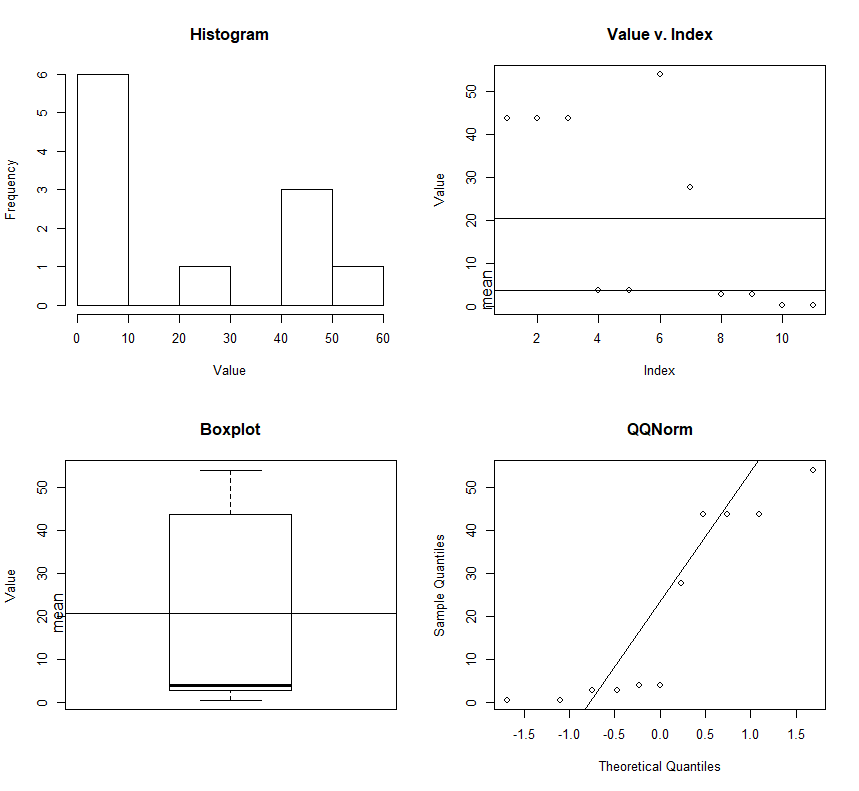


11 CUCULIFORMES

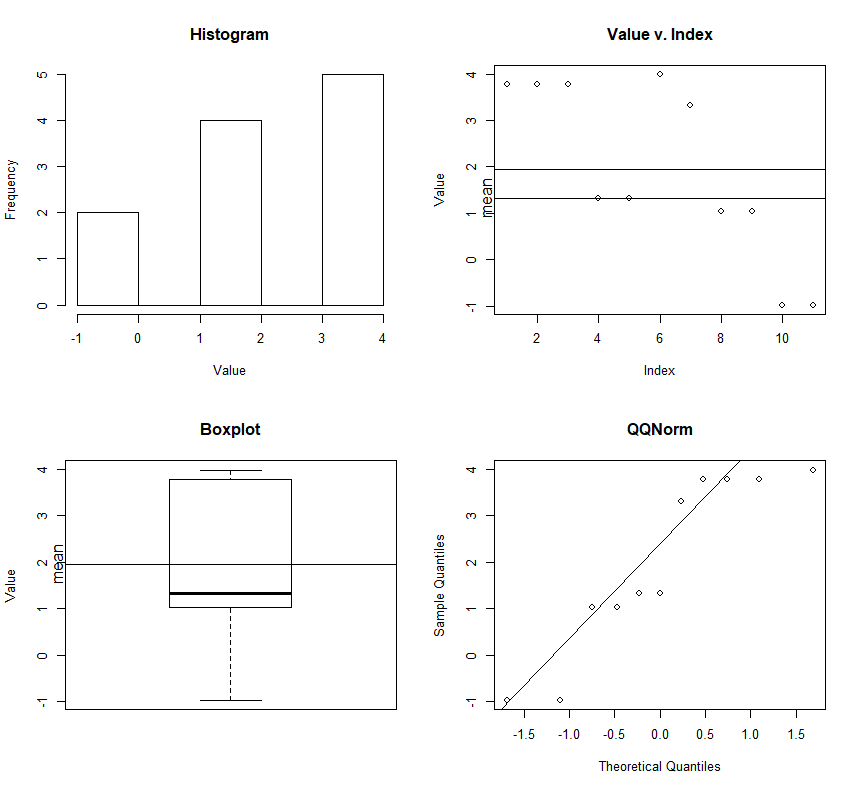
> summary(rttree[rttree$ORDER=="CUCULIFORMES",4]);eda.wp((rttree[rttree$ORDER=="CUCULIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.3786 2.8150 3.7590 20.6000 43.6900 53.9600



[APP: Yo! A weird biphasic thing. Here is log transformed.

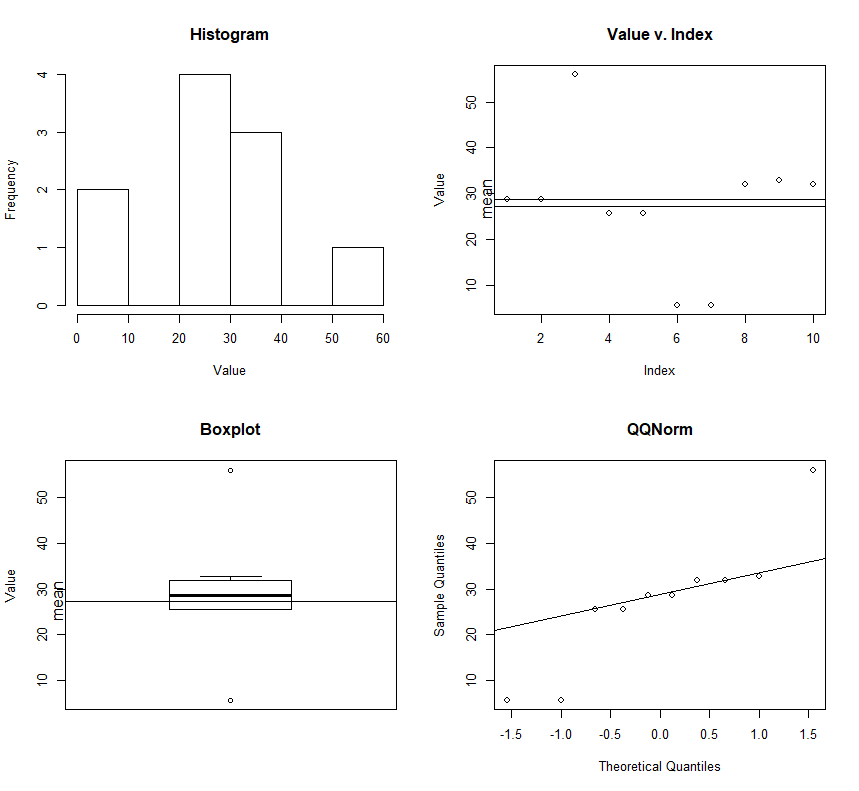


10 PROCELLARIIFORMES

> summary(rttree[rttree$ORDER=="PROCELLARIIFORMES",4]);eda.wp((rttree[rttree$ORDER=="PROCELLARIIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

5.689 25.620 28.670 27.260 31.950 55.990



[APP: appears tightly constrained, but this may be in large part a small sample size effect?

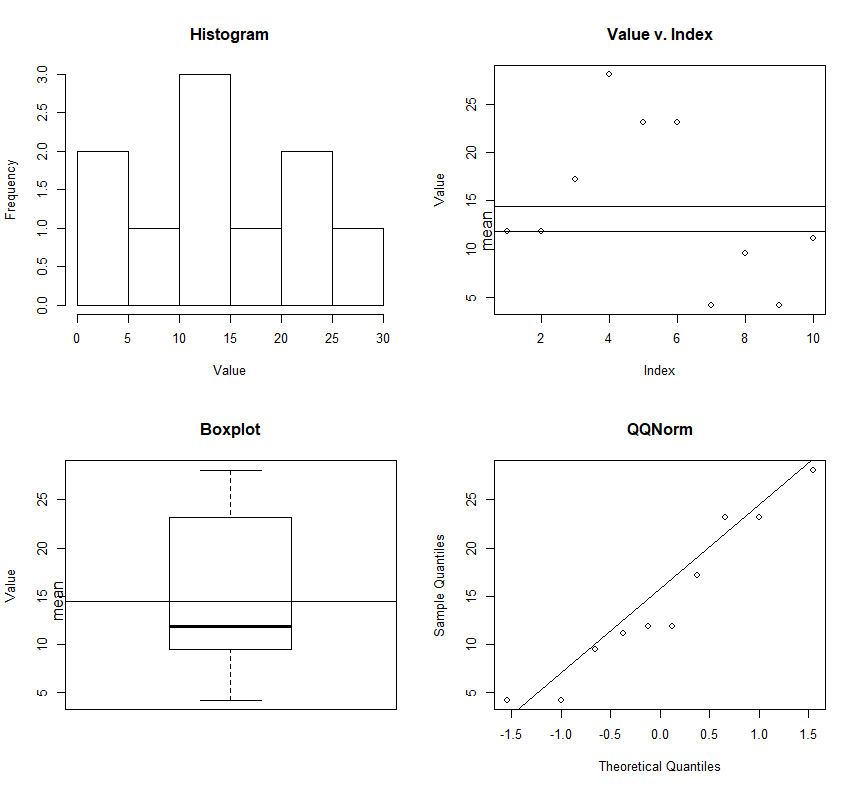
Or I wonder if ease of availability of samples has some possible bias?]

10 FALCONIFORMES

> summary(rttree[rttree$ORDER=="FALCONIFORMES",4]);eda.wp((rttree[rttree$ORDER=="FALCONIFORMES",4]))

Min. 1st Qu. Median Mean 3rd Qu. Max.

4.205 9.905 11.860 14.440 21.660 28.100



Notes:

Many questions raised. A reasonable number of values certainly need confirmation, which may also reveal some data manipulation/editing problem.

The fundamental INTERESTIND question to ME is: Is it reasonable, or possible to create an expectation regarding the distribution of terminal branch lengths for some partitioned set of AVES.

My anterior expectation is (was) YES, it should be on a broad scale, and that further it would be reasonable to expect some predictable differences. Small birds with shorter generation times would, I expect, have shorter terminal branch lengths in settings where multiple selective pressures would make faster divergence/speciation possible.

Some may well defy the body size- generation time effect.. ? Mousebirds?

What partitioning level (ORDER, Family, Genus) would be the most instructive for investigation. – this of course partly depends on data/sample size.

Are geographic difference discernable? Africa – looks like S.Am. but without the Andes. How, for example would they compare.

If the distributions demonstrated hold up as repeatable when this is repeated and more formally done, what is implied by the distributions (to me this is THE interesting question, as it implies a great deal about what KINDS of statistical and analytic techniques are appropriate and which are not.

Is log transformation appropriate, and if so when, and why. This depends in part on what “terminal branch length” really “measures”. If it couples with “time”, then I would expect log transformation to have some support.

It certainly looks like using the mean (average) to take an expectation is pretty clearly a bad idea, as assumption of Gaussian character is often substantially violated, other times the data look “well behaved” without transformation… what gives with that?

If one thinks of Branch length as an indicator of TIME, then branch length is essentially a “time-until”, or “time-since” measure. Time-from-branching-until-specimen-capture, or time-since-branching-until-speicmen-capture.

Statistical analyses of such data are almost ALWAYS done inappropriately, and a suite of tools (largely unknown or ignored in most of biology) has grown up to analyze such data – “Survival analysis” developed first in the industrial world, but subsequently developed and used extensively in medicine, for clinical trials &c.