Detailed description of Statistical Analysis Methods

* Here are the Questions:
  + **Can you predict fresh weight from the box the pad fits in (HWT) and the width of the circle with the same perimeter as the pad (D)?**

**Initial Data Investigations**

The initial data were taken from the field records of cactus pad dimensions as measured destructively in the field, with five measures for each pad (n = 448) organized into 14 accessions. These measures consisted of height (the maximum length of the pad), width (maximum length of line segment perpendicular to pad), thickness, and diameter (derived from the perimeter of the pad), as well as the fresh weight of the pad. To phrase these data geometrically, each pad \_\_\_\_\_\_ the smallest box it could fit in and the diameter of the circle which had its perimeter (Table \_\_), and the weight of the pad.

*Normality*

The first data processing step was an expansion of the dataset to derive additional measures used to assess the distributions of the data. The ratios of fresh weight to the other four initial measures were calculated, as well as the ratio of diameter to width, and the dry weights of each of the pads (calculated from the average reduction in weight between the fresh and dry pads for each accession). From extensive Q-Q plot generation on both an overall and per-accession basis for all measures thus far, it was determined that, due to consistently observed heavy tails in various distributions, subsequent analyses must be conducted with the assumption of non-normal data.

*Interaccessional Variance in Distributions*

Subsequently, the Tukey tests were employed to determine the range of variance for each measure, and in the process gain further insight towards measure interrelation. These Tukey tests were also represented as non-simple hypergraphs for ease of visual comparison. The tests demonstrated that, while some accessions had significant differences in the distribution of their means in some measures but not others, showing that significant morphological differences between accessions can be quantified to a reasonable degree even with a minimal sample size.

*Measure-Measure Intercorrelations*

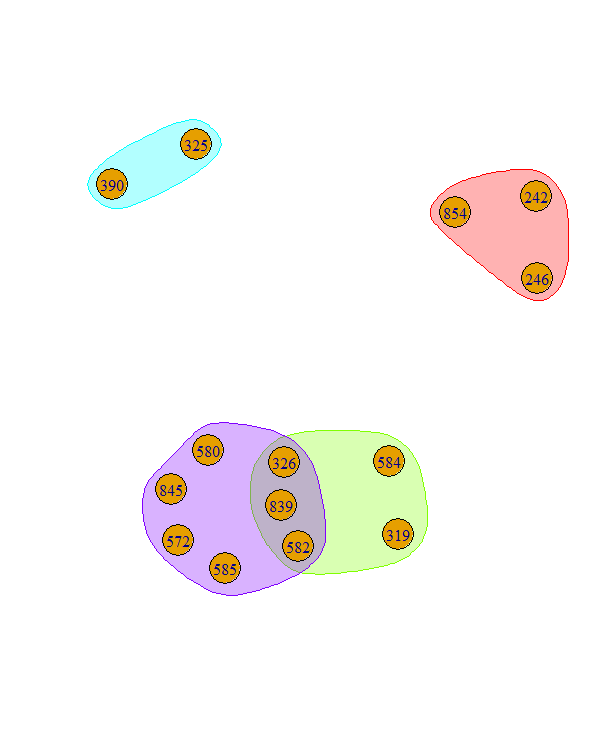
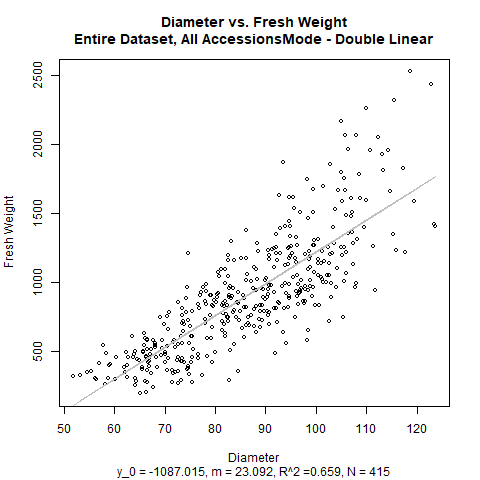
The first explicit assessments of intermeasure correlation were conducted over the five base measures and a Diameter:Width ratio measure. The latter was included for the purposes of outlier selection – extreme D:W ratios were correlated with pads that had extreme or nonsensical data (ex. Width was larger than Height, or two pads with otherwise similar characteristics had drastically different heights). These measure intercorrelations yielded several significant relationships:

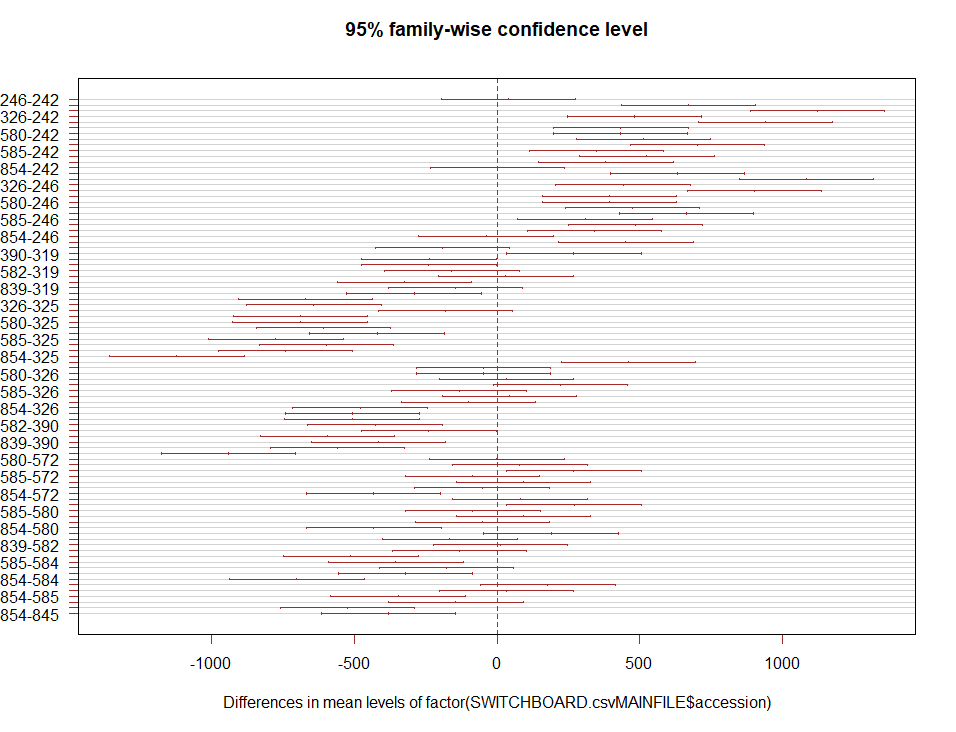


*Simple Linear Modeling*

*Ellipse Modeling*

* Start with initial data
  + completed
  + fresh\_weight
  + width
  + height
  + diameter
  + thickness
  + accession
  + species
* Shape of the data
  + Predominantly non-normal – heavy tails on one side or another.
  + Further statistical comparisons require non-normal comparisons 🡪 Tukey
* Intermeasure analysis
  + Measures must necessarily be interrelated –
    - Height and width not great predictors by themselves, but H\*W as a proxy for area has a stronger relation to fresh weight given a fairly uniform thickness
  + Hence the extent of measure interrelation needs to be determined to make accurate models
  + Log/log and linear plots
  + Tukey analysis to demonstrate that some accession have demonstrable variation in the measures, rather than all existing on a single spectrum.
* Linear Modeling R^2 and SBC
  + How far can we get with straightforward linear models? Using SBC/AIC (harshest criterion), we actually do pretty well!
  + The fact that accessions differ in their particular ranges for each measure, and that individual fits and general fits both do well overall (verify!) indicate that it may be possible to construct a coefficientless formula relying only on the intrinsic measures.
* Elliptical modeling approach





AdjR^2 100



AdjR^2 95



AdjR^2 90



SBC 100



SBC 95



SBC 90

