The data set is available as SQLite database and csv file. There are X measurements for X organisms of X variables. Some data entries are missing (X total, across the data set).

Caveats:

Data sets are not orthogonal and not all organisms are equally represented.

Check for collinearity

**Piecewise regression** was applied [fig 1.]

Diagnose collinear variables

**Multiple regression** and **analysis of covariance** [fig 3 and S3].

The maximal model can be described with a function: y=β0+ β1x1+…+ β13x13

y – meiotic metaphase spindle length (pole to pole or aster to aster).

To determine, which variables should be used

(forward propagation)

Tree model accounts for approximately 90% of the variation in mitotic spindle length, out of which.

Given that the model explains 85% of variance, we used the

Unsupervised learning and logistic regression [fig 4]

We applied logistic regression using meiotic/mitotic

Shapiro test for normality for continuous measurement

Low rank matrix approximation was used to

Data was scaled

Kmeans clustering and GMM

Chi-square

Histograms for the groups

Python and R code used to process, analyze and plot data is available on GitHub <https://github.com/emmaggie/SPINDLE_ZOO> (academic license to M.S.).

The IPython notebooks can be viewed with Notebook Viewer (<http://nbviewer.ipython.org/>).