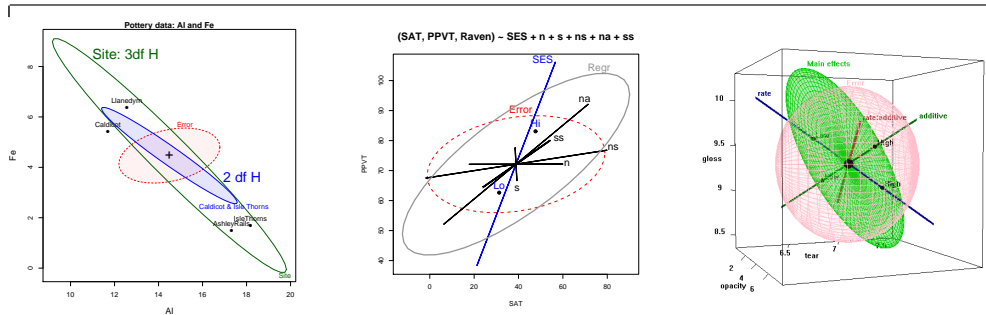


Recent Advances in Visualizing Multivariate Linear Models

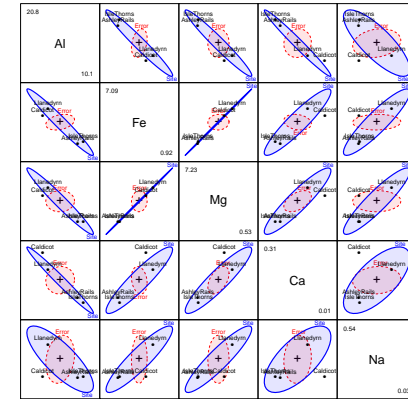
Michael Friendly Matthew Sigal

Statistics Day @ York, April 5, 2013



Outline

- 1 Background
 - Visual overview
 - Data ellipses
 - The Multivariate Linear Model
 - Motivating example
- 2 Hypothesis-Error (HE) plots
 - Visualizing H and E (co)variation
 - MANOVA designs
 - MREG designs
- 3 Reduced-rank displays
 - Low-D displays of high-D data
 - Canonical discriminant HE plots
- 4 Recent extensions
 - Canonical correlation
 - Robust MLMs
 - Influence diagnostics for MLMs
- 5 Conclusions



Slides: <http://datavis.ca/papers/ssc2013/>

Background

Background

Visual overview

Precepts of this work

Visualization

Should be fundamental in statistical theory & practice.

"If I can't picture it, I can't understand it." — Albert Einstein

*"In certain problems it was necessary to develop the **picture as the method** before the mathematics could be really done"* — Richard Feynman

Theory into Practice

*"The **practical power** of a statistical test is the product of its' statistical power and the probability of use."* — J. W. Tukey, 1959

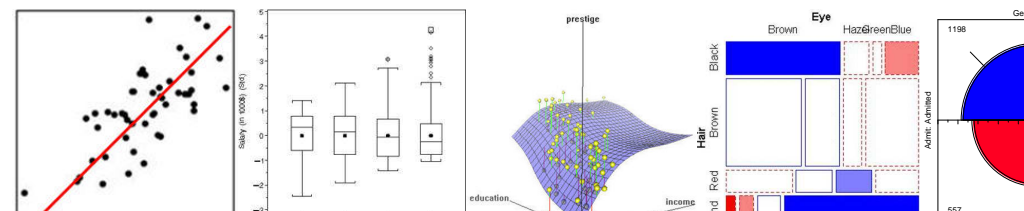
Computation and Implementation

- Modern statistical methods are often mathematically complex and computationally intensive (e.g., bootstrap, MCMC, asymptotics)
- A general implementation allows these to be tested **studied as statistical objects** and **find flaws** in theory or implementation.

Introduction: The LM family and friends

Models, graphical methods and opportunities

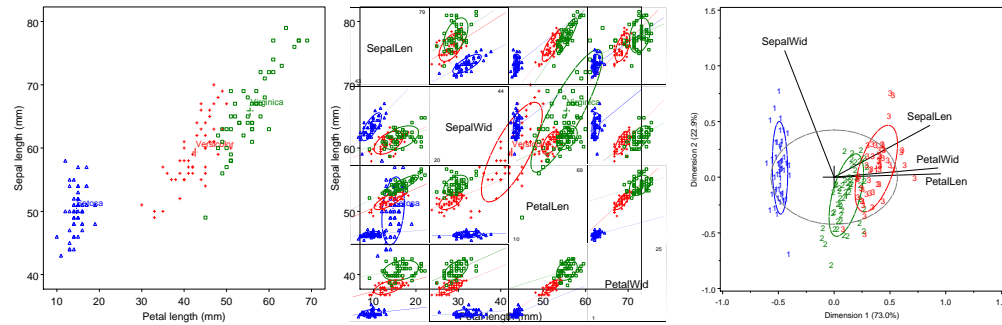
	Classical linear models	Generalized linear models	
# of response variables	1 LM family: $E(y)=X\beta$, $V(y X)=\sigma^2I$ ANOVA, regression, ... Many graphical methods: effect plots, spread-leverage, influence, ...	GLM: $E(y)=g^{-1}(X\beta)$, $V=V[g^{-1}(X\beta)]$ poisson, logistic, loglinear, ... Some graphical methods: mosaic plots, 4fold plots, diagnostic plots, ...	# of response variables
	2+ MLM: $E(Y)=X\beta$, $V(Y X)=I\otimes\Sigma$ MANOVA, MReg, ... Graphical methods: ???	MGLM: ??? Graphical methods: ???	



Visual overview: Multivariate data, $\mathbf{Y}_{n \times p}$

What we know how to do well (almost)

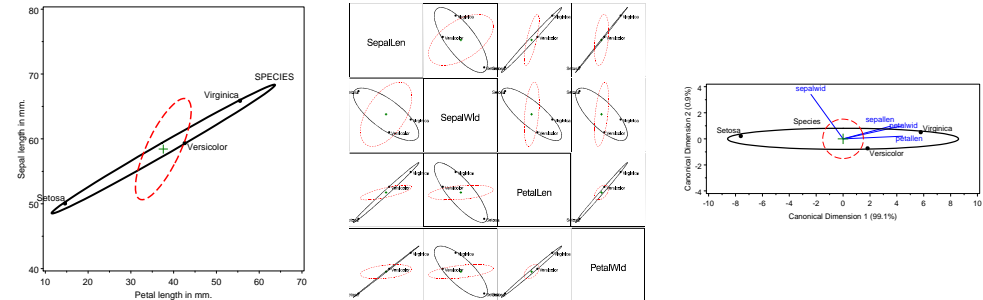
- 2 vars: Scatterplot + annotations (data ellipses)
- p vars: Scatterplot matrix (all pairs)
- p vars: Reduced-rank display— show max. total variation \mapsto biplot



Visual overview: Multivariate linear model, $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}$

What is new here?

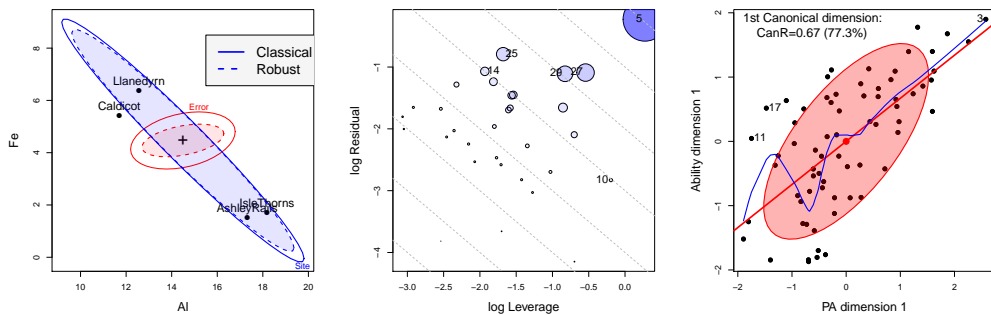
- 2 vars: HE plot— data ellipses of \mathbf{H} (fitted) and \mathbf{E} (residual) SSP matrices
- p vars: HE plot matrix (all pairs)
- p vars: Reduced-rank display— show max. \mathbf{H} wrt. $\mathbf{E} \mapsto$ Canonical HE plot



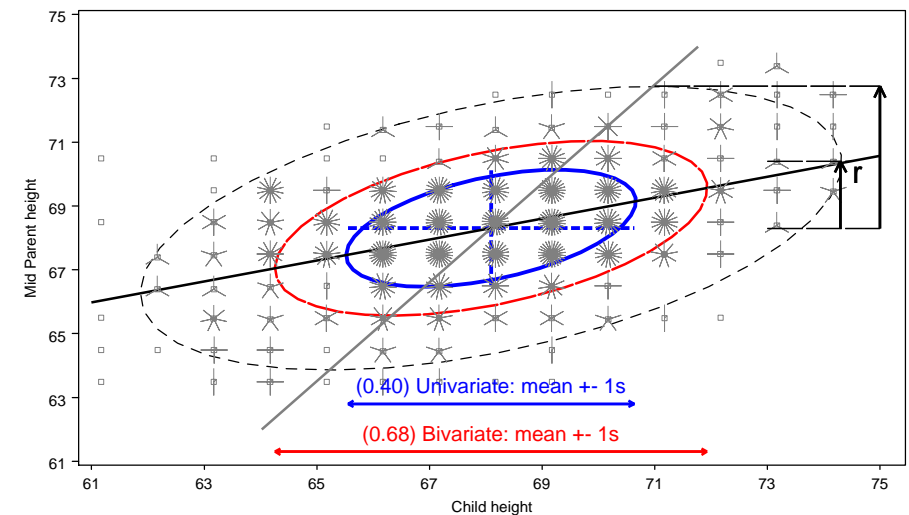
Visual overview: Recent extensions

Extending univariate methods to MLMs:

- Robust estimation for MLMs
- Influence measures and diagnostic plots for MLMs
- Visualizing canonical correlation analysis



Data Ellipses: Galton's data



Galton's data on Parent & Child height: 40%, 68% and 95% data ellipses

The Data Ellipse: Details

• Visual summary for bivariate relations

- **Shows:** means, standard deviations, correlation, regression line(s)
- **Defined:** set of points whose squared Mahalanobis distance $\leq c^2$,

$$D^2(\mathbf{y}) \equiv (\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{S}^{-1} (\mathbf{y} - \bar{\mathbf{y}}) \leq c^2$$

\mathbf{S} = sample variance-covariance matrix

- **Radius:** when \mathbf{y} is \approx bivariate normal, $D^2(\mathbf{y})$ has a large-sample χ^2_2 distribution with 2 degrees of freedom.
 - $c^2 = \chi^2_2(0.40) \approx 1$: 1 std. dev univariate ellipse– 1D shadows: $\bar{y} \pm 1s$
 - $c^2 = \chi^2_2(0.68) \approx 2.28$: 1 std. dev bivariate ellipse
 - $c^2 = \chi^2_2(0.95) \approx 6$: 95% data ellipse, 1D shadows: Scheffé intervals
- **Construction:** Transform the unit circle, $\mathcal{U} = (\sin \theta, \cos \theta)$,

$$\mathcal{E}_c = \bar{\mathbf{y}} + c\mathbf{S}^{1/2}\mathcal{U}$$

$\mathbf{S}^{1/2}$ = any “square root” of \mathbf{S} (e.g., Cholesky)

- **Robustify:** Use robust estimate of \mathbf{S} , e.g., MVE (mimimum volume ellipsoid)
- **p variables:** Extends naturally to p -dimensional ellipsoids

The univariate linear model

- **Model:** $\mathbf{y}_{n \times 1} = \mathbf{X}_{n \times q} \boldsymbol{\beta}_{q \times 1} + \boldsymbol{\epsilon}_{n \times 1}$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$
- **LS estimates:** $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- **General Linear Test:** $H_0 : \mathbf{C}_{h \times q} \boldsymbol{\beta}_{q \times 1} = \mathbf{0}$, where \mathbf{C} = matrix of constants; rows specify h linear combinations or contrasts of parameters.
- e.g., Test of $H_0 : \beta_1 = \beta_2 = 0$ in model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$

$$\mathbf{C}\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- All \rightarrow F-test: How big is SS_H relative to SS_E ?

$$F = \frac{SS_H/\text{df}_h}{SS_E/\text{df}_e} = \frac{MS_H}{MS_E} \rightarrow (MS_H - F MS_E) = 0$$

The multivariate linear model

- **Model:** $\mathbf{Y}_{n \times p} = \mathbf{X}_{n \times q} \mathbf{B}_{q \times p} + \mathbf{U}$, for p responses, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p)$
- **General Linear Test:** $H_0 : \mathbf{C}_{h \times q} \mathbf{B}_{q \times p} = \mathbf{0}_{h \times p}$
- Analogs of sums of squares, SS_H and SS_E are $(p \times p)$ matrices, \mathbf{H} and \mathbf{E} ,

$$\mathbf{H} = (\mathbf{C}\hat{\mathbf{B}})^T [\mathbf{C}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T]^{-1} (\mathbf{C}\hat{\mathbf{B}}),$$

$$\mathbf{E} = \mathbf{U}^T \mathbf{U} = \mathbf{Y}^T [\mathbf{I} - \mathbf{H}] \mathbf{Y}.$$

- Analog of univariate F is

$$\det(\mathbf{H} - \lambda \mathbf{E}) = 0,$$

- How big is \mathbf{H} relative to \mathbf{E} ?
 - Latent roots $\lambda_1, \lambda_2, \dots, \lambda_s$ measure the “size” of \mathbf{H} relative to \mathbf{E} in $s = \min(p, \text{df}_h)$ orthogonal directions.
 - Test statistics (Wilks' Λ , Pillai trace criterion, Hotelling-Lawley trace criterion, Roy's maximum root) all combine info across these dimensions

Motivating Example: Romano-British Pottery

Tubb, Parker & Nicholson analyzed the chemical composition of 26 samples of Romano-British pottery found at four kiln sites in Britain.

- Sites: Ashley Rails, Caldicot, Isle of Thorns, Llanedryn
- Variables: aluminum (Al), iron (Fe), magnesium (Mg), calcium (Ca) and sodium (Na)
- \rightarrow One-way MANOVA design, 4 groups, 5 responses

```
R> library(heplots)
```

```
R> Pottery
```

	Site	Al	Fe	Mg	Ca	Na
1	Llanedryn	14.4	7.00	4.30	0.15	0.51
2	Llanedryn	13.8	7.08	3.43	0.12	0.17
3	Llanedryn	14.6	7.09	3.88	0.13	0.20
	...					
25	AshleyRails	14.8	2.74	0.67	0.03	0.05
26	AshleyRails	19.1	1.64	0.60	0.10	0.03

Motivating Example: Romano-British Pottery

Questions:

- **Can** the content of Al, Fe, Mg, Ca and Na differentiate the sites?
- **How to understand** the contributions of chemical elements to discrimination?

Numerical answers:

```
R> pottery.mod <- lm(cbind(Al, Fe, Mg, Ca, Na) ~ Site)
R> Manova(pottery.mod)
```

Type II MANOVA Tests: Pillai test statistic

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Site	3		1.55		4.30		15		60	2.4e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

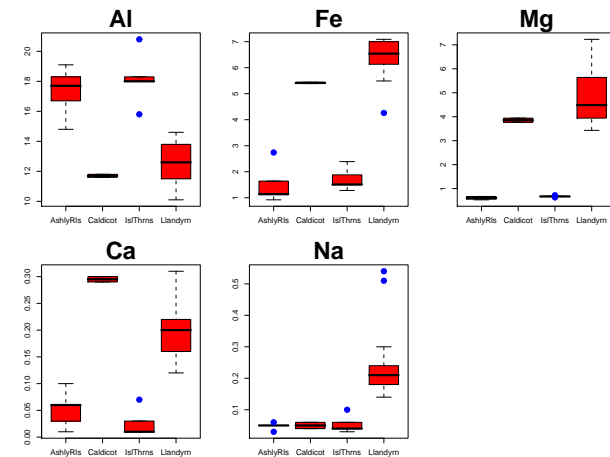
What have we learned?

- **Can**: YES! We can discriminate sites.
- But: **How to understand** the pattern(s) of group differences: ???

Motivating Example: Romano-British Pottery

Univariate plots are limited

- Do not show the *relations* of variables to each other

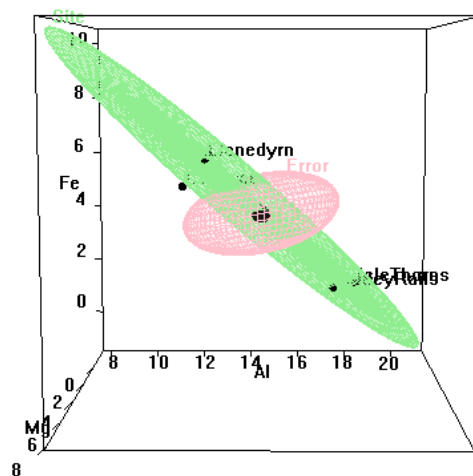


Motivating Example: Romano-British Pottery

Visual answer: HE plot

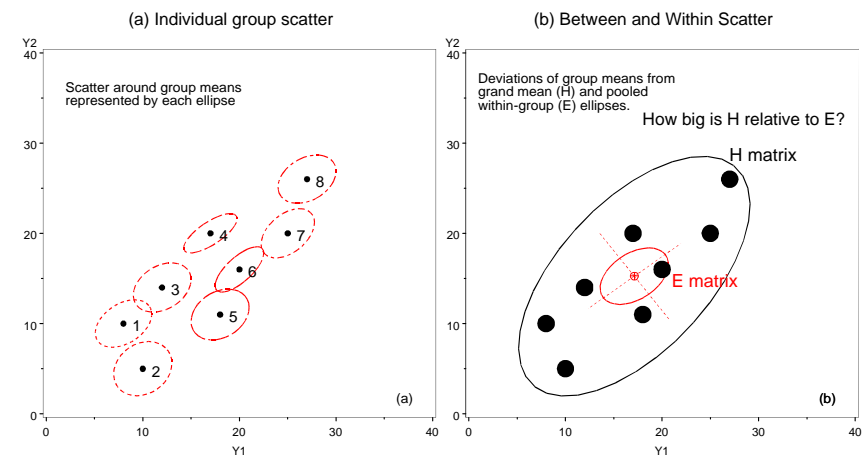
- Shows variation of means (**H**) relative to residual (**E**) variation
- Size and orientation of **H** wrt **E**: *how much* and *how* variables contribute to discrimination
- Evidence scaling: **H** is scaled so that it projects outside **E** iff null hypothesis is rejected.

Run heplot-movie.ppt



```
R> heplot3d(pottery.mod)
```

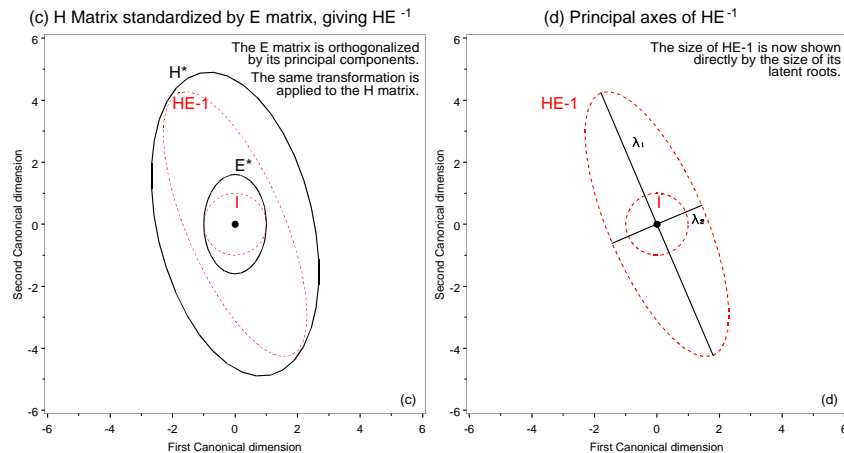
HE plots: Visualizing H and E (co) variation



Ideas behind multivariate tests: (a) Data ellipses; (b) **H** and **E** matrices

- **H** ellipse: data ellipse for fitted values, $\hat{y}_{ij} = \bar{y}_j$.
- **E** ellipse: data ellipse of residuals, $\hat{y}_{ij} - \bar{y}_j$.

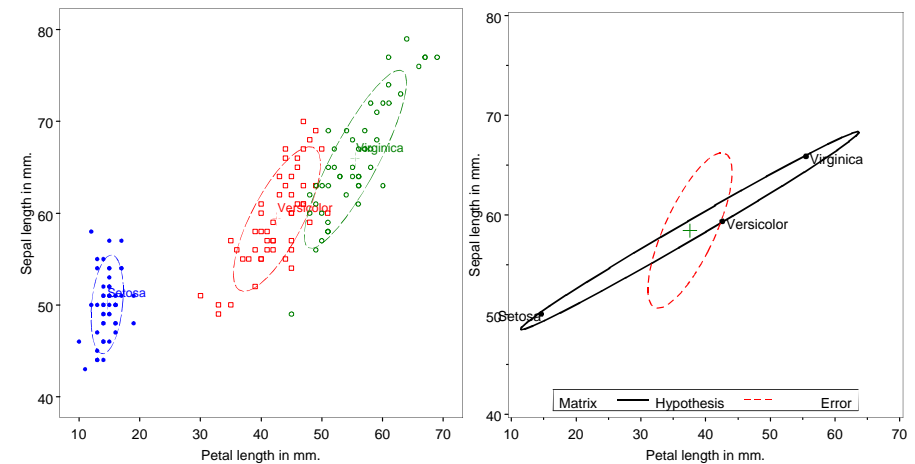
HE plots: Visualizing multivariate hypothesis tests



Ideas behind multivariate tests: latent roots & vectors of \mathbf{HE}^{-1}

- $\lambda_i, i = 1, \dots, df_h$ show size(s) of \mathbf{H} relative to \mathbf{E} .
- latent vectors show canonical directions of maximal difference.

HE plot for iris data



(a) Data ellipses and (b) \mathbf{H} and \mathbf{E} matrices (scaled by $1/df_e$: effect size)

- \mathbf{H} ellipse: data ellipse for fitted values, $\hat{\mathbf{y}}_{ij} = \bar{\mathbf{y}}_j$.
- \mathbf{E} ellipse: data ellipse of residuals, $\hat{\mathbf{y}}_{ij} - \bar{\mathbf{y}}_j$.

HE plot details: \mathbf{H} and \mathbf{E} matrices

Recall the data on 5 chemical elements in samples of Romano-British pottery from 4 kiln sites:

```
R> summary(Manova(pottery.mod))
```

Sum of squares and products for error:

	Al	Fe	Mg	Ca	Na
Al	48.29	7.080	0.608	0.106	0.589
Fe	7.08	10.951	0.527	-0.155	0.067
Mg	0.61	0.527	15.430	0.435	0.028
Ca	0.11	-0.155	0.435	0.051	0.010
Na	0.59	0.067	0.028	0.010	0.199

Term: Site

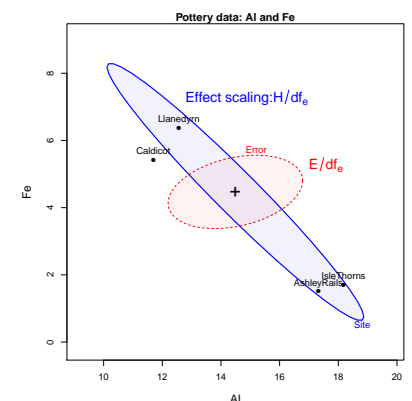
Sum of squares and products for hypothesis:

	Al	Fe	Mg	Ca	Na
Al	175.6	-149.3	-130.8	-5.89	-5.37
Fe	-149.3	134.2	117.7	4.82	5.33
Mg	-130.8	117.7	103.4	4.21	4.71
Ca	-5.9	4.8	4.2	0.20	0.15
Na	-5.4	5.3	4.7	0.15	0.26

- \mathbf{E} matrix: Within-group (co)variation of residuals
 - diag: SSE for each variable
 - off-diag: \sim partial correlations
- \mathbf{H} matrix: Between-group (co)variation of means
 - diag: SSH for each variable
 - off-diag: \sim correlations of means
- How big is \mathbf{H} relative to \mathbf{E} ?
- Ellipsoids: $\dim(\mathbf{H}) = \text{rank}(\mathbf{H}) = \min(p, df_h)$

HE plot details: Scaling \mathbf{H} and \mathbf{E}

- The \mathbf{E} ellipse is divided by $df_e = (n - p) \rightarrow$ data ellipse of residuals
 - Centered at grand means \rightarrow show factor means in same plot.
- "Effect size" scaling- $\mathbf{H}/df_e \rightarrow$ data ellipse of fitted values.
- "Significance" scaling- \mathbf{H} ellipse protrudes beyond \mathbf{E} ellipse iff H_0 can be rejected by Roy maximum root test
 - $H/(\lambda_\alpha df_e)$ where λ_α is critical value of Roy's statistic at level α .
 - direction of \mathbf{H} wrt $\mathbf{E} \mapsto$ linear combinations that depart from H_0 .

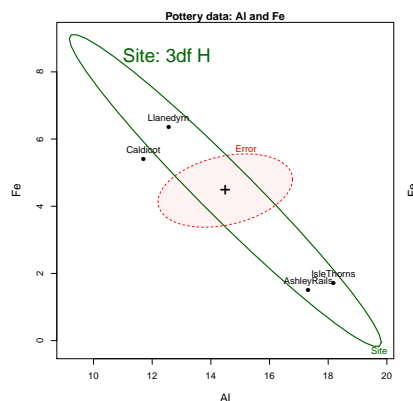


```
R> heplot(pottery.mod, size="effect")
size="evidence")
```

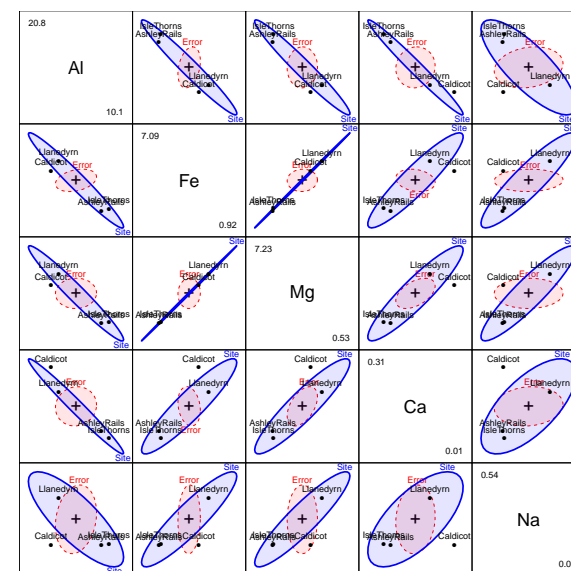
```
R> heplot(pottery.mod,
```

HE plot details: Contrasts and linear hypotheses

- An overall effect \mapsto an **H** ellipsoid of $s = \min(p, df_h)$ dimensions
- Linear hypotheses, of the form $H_0 : \mathbf{C}_{h \times q} \mathbf{B}_{q \times p} = \mathbf{0}_{h \times p} \mapsto$ sub-ellipsoid of dimension h
- 1D tests and contrasts \mapsto degenerate 1D ellipses (lines)



HE plot matrices: All bivariate views



AL stands out –
opposite pattern
 $r(\overline{Fe}, \overline{Mg}) \approx 1$

► Jump to low-D

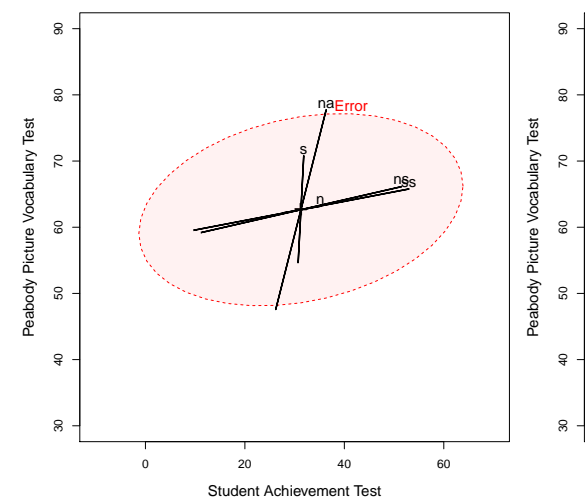
R> pairs(pottery.mod)

HE plots for Multivariate Multiple Regression

- **Model:** $\mathbf{Y} = \mathbf{XB} + \mathbf{U}$, where cols of \mathbf{X} are quantitative.
- **Overall test:** $H_0 : \mathbf{B} = \mathbf{0}$ (all coefficients for all responses are zero)
 - $\rightarrow \mathbf{C} = \mathbf{I}$ in GLT $\rightarrow \mathbf{H} = \hat{\mathbf{B}}^T (\mathbf{X}^T \mathbf{X})^{-1} \hat{\mathbf{B}} = \hat{\mathbf{Y}}^T \hat{\mathbf{Y}}$
- **Individual predictors:** $H_0 : \beta_i = 0$
 - $\rightarrow \mathbf{C} = (0, 0, \dots, 1, 0, \dots, 0) \rightarrow \mathbf{H}_i = \hat{\beta}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \hat{\beta}_i$
- **HE plot**
 - Overall **H** ellipse: how predictors relate collectively to responses
 - Individual **H** ellipses ($\text{rank}(\mathbf{H})=1 \rightarrow$ vectors):
 - orientation \rightarrow relation of \mathbf{x}_i to $\mathbf{y}_1, \mathbf{y}_2$
 - length \rightarrow strength of relation
 - collection of individual **H** vectors \rightarrow how predictors contribute to overall test.

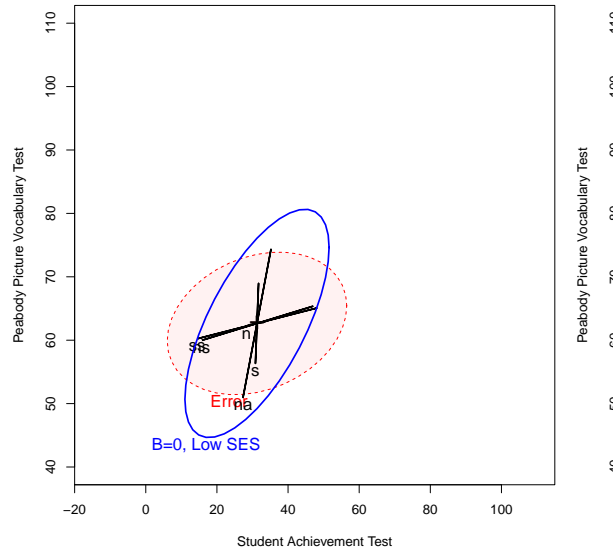
HE plots for MMRA: Example

- Rohwer data on $n = 37$ low SES children, for 5 PA tasks (N, S, NS, NA, SS) predicting intelligence/achievement (PPVT, SAT, Raven)
- Only NA is individually significant (in this view)
- ... but overall test highly significant
- NA & S contribute to predicting PPVT
- NS & SS contribute to predicting SAT



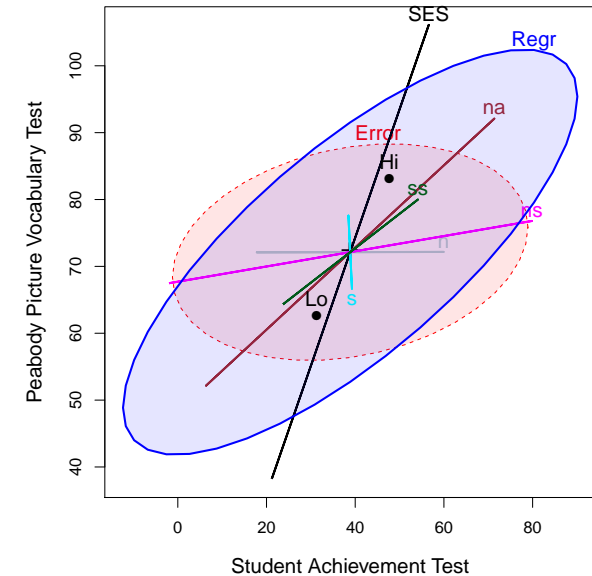
HE plots for MMRA: MANCOVA

- Rohwer data on $n_1 = 37$ low SES, and $n_2 = 32$ high SES children
- Fit separate regressions for each group
- Are regressions parallel?
- Are they coincident?



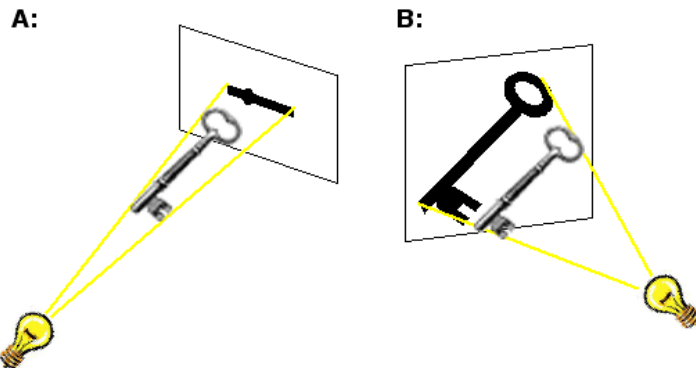
HE plots for MMRA: MANCOVA

- Rohwer data on $n_1 = 37$ low SES, and $n_2 = 32$ high SES children
- Fit MANCOVA model (assuming equal slopes)



Low-D displays of high-D data

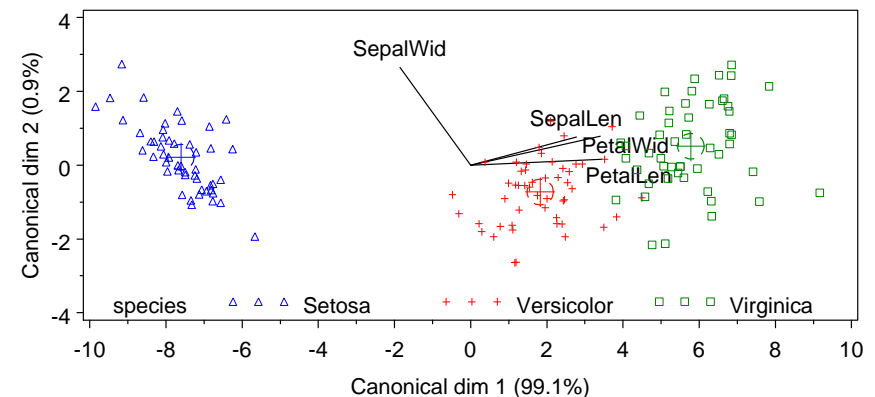
- High-D data often shown in 2D (or 3D) views— orthogonal projections in variable space— **scatterplot**
- Dimension-reduction** techniques: project the data into subspace that has the largest *shadow*— e.g., accounts for largest variance.
- low-D approximation to high-D data



A: minimum-variance projection; B: maximum variance projection

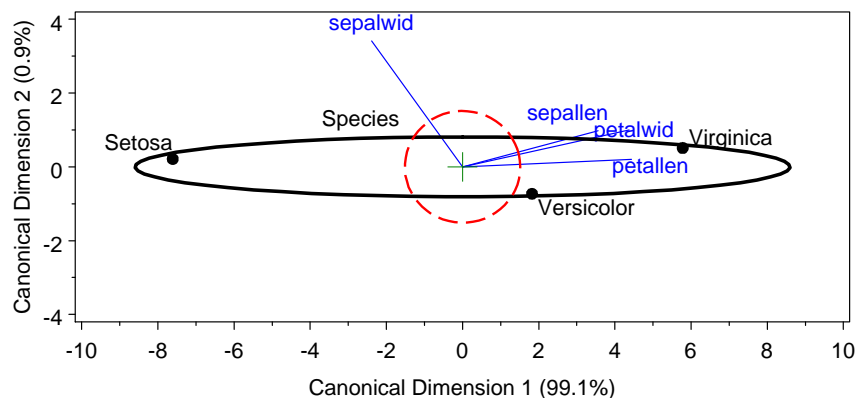
Canonical discriminant HE plots

- As with biplot, we can visualize MLM hypothesis variation for *all* responses by projecting \mathbf{H} and \mathbf{E} into low-rank space.
- Canonical projection:** $\mathbf{Y}_{n \times p} \mapsto \mathbf{Z}_{n \times s} = \mathbf{Y}\mathbf{E}^{-1/2}\mathbf{V}$, where \mathbf{V} = eigenvectors of $\mathbf{H}\mathbf{E}^{-1}$.
- This is the view that maximally discriminates among groups, ie max. \mathbf{H} wrt \mathbf{E} !



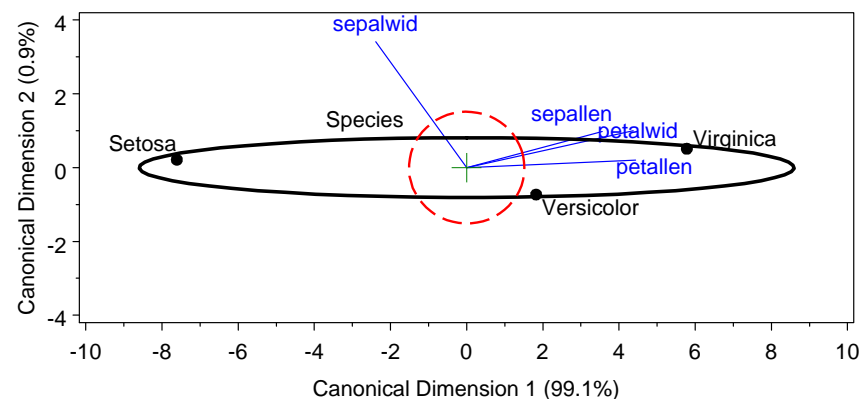
Canonical discriminant HE plots

- Canonical HE plot is just the HE plot of canonical scores, (z_1, z_2) in 2D,
- or, z_1, z_2, z_3 , in 3D.
- As in biplot, we add vectors to show relations of the y_i response variables to the canonical variates.
- variable vectors here are **structure coefficients** = correlations of variables with canonical scores.



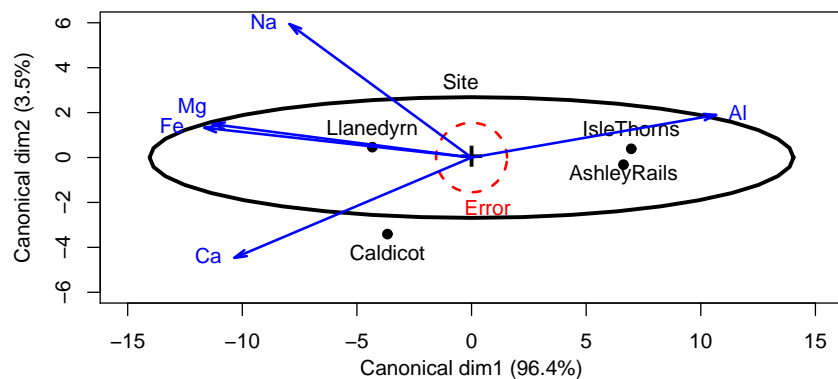
Canonical discriminant HE plots: Properties

- Canonical variates are uncorrelated: **E** ellipse is spherical
- \mapsto axes must be equated to preserve geometry
- Variable vectors show how variables discriminate among groups
- Lengths of variable vectors \sim contribution to discrimination



Canonical discriminant HE plots: Pottery data

- Canonical HE plots provide 2D (3D) visual summary of **H** vs. **E** variation
- Pottery data: $p = 5$ variables, 4 groups $\mapsto df_H = 3$
- Variable vectors: Fe, Mg and Al contribute to distinguishing (Caldicot, Llandryn) from (Isle Thorns, Ashley Rails): 96.4% of mean variation
- Na and Ca contribute an additional 3.5%. **End of story!**

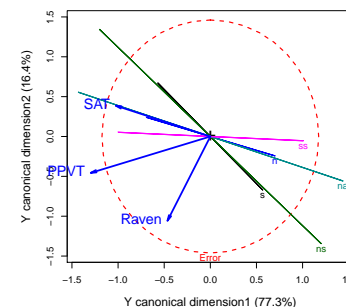
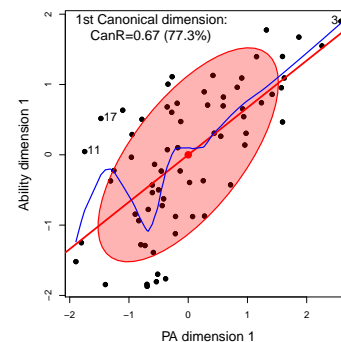


Visualizing Canonical Correlation Analysis

- Basic idea: another instance of low-rank approximation

CCA is to *MMReg* as *CDA* is to *MANOVA*

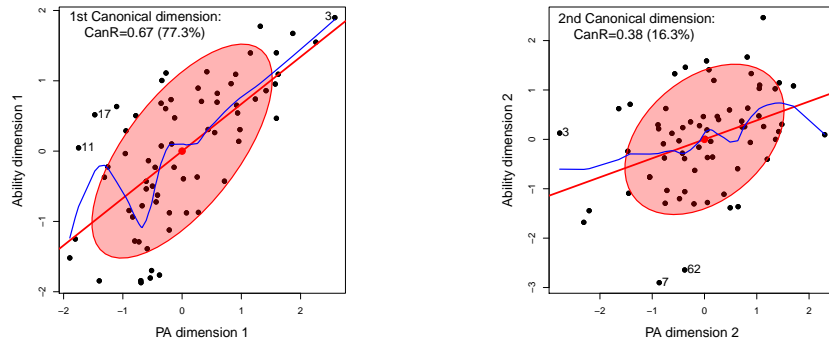
- \rightarrow For quantitative predictors, provides an alternative view of $\mathbf{Y} \sim \mathbf{XB}$ in space of maximal (canonical) correlations.
- The candisc package implements two new views for CCA:
 - `plot()` method to show canonical (X, Y) variates as **data**
 - `heplot()` method to show (X, Y) relations as **heplots** for **Y** in CAN space.



CCA Example: Rohwer data, Ability and PA tests

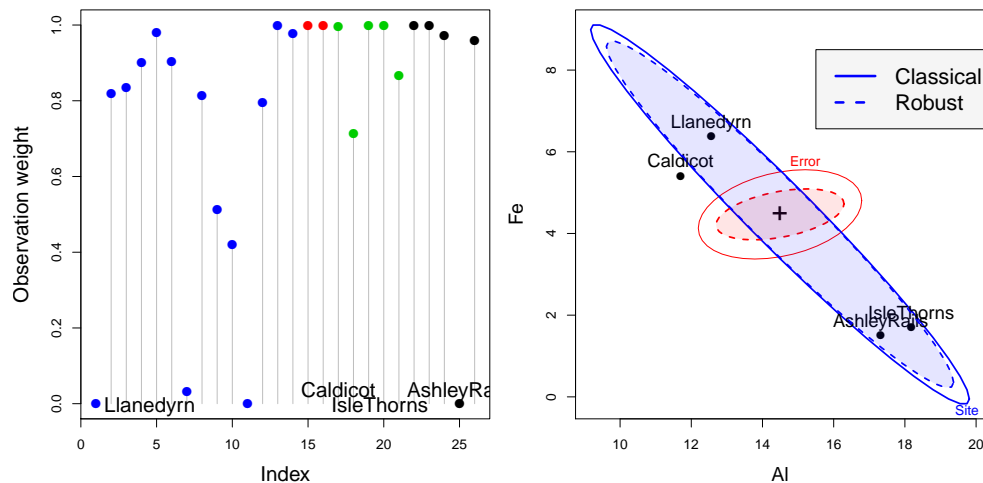
- `plot()` method shows canonical variates for **X** and **Y** on one dimension
- Smoother shows possible non-linearity
- Point identification highlights unusual observations

```
R> library(candisc)
R> cc <- cancor(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss,
+             data=Rohwer, set.names=c("PA", "Ability"))
R> plot(cc, smooth=TRUE, id.n=3)
R> plot(cc, smooth=TRUE, id.n=3, which=2)
```



Robust MLMs: Example

For the Pottery data:



- Some observations are given weights ~ 0
- The **E** ellipse is considerably reduced, enhancing apparent significance

Robust MLMs

- R has a large collection of packages dealing with robust estimation:
 - `robust::lmrob()`, `MASS::rlm()`, for univariate LMs
 - `robust::glmrob()` for univariate *generalized* LMs
 - **High breakdown-bound** methods for robust PCA and robust covariance estimation
 - However, none of these handle the **fully general MLM**
- The `heplots` package now provides `robmlm()` for robust MLMs:
 - Uses a simple M-estimator via iteratively re-weighted LS.
 - Weights: calculated from Mahalanobis squared distances, using a simple robust covariance estimator, `MASS::cov.trob()` and a weight function, $\psi(D^2)$.

$$D^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^T \mathbf{S}_{\text{trob}}^{-1} (\mathbf{Y} - \hat{\mathbf{Y}}) \sim \chi_p^2 \quad (1)$$

- This fully extends the "mlm" class
- Compatible with other mlm extensions: `car::Anova` and `heplots::heplot`.
- Downside: Does not incorporate modern consistency factors or other niceties.

Influence diagnostics for MLMs

- Influence measures and diagnostic plots are well-developed for univariate LMs
 - Influence measures: Cook's D, DFFITS, dfbetas, etc.
 - Diagnostic plots: Index plots, `car::influencePlot()` for LMs
 - However, these have been unavailable for MLMs
- The `mvinfluence` package now provides:
 - Calculation for multivariate analogs of univariate influence measures (following Barrett & Ling, 1992), e.g., Hat values & Cook's D:

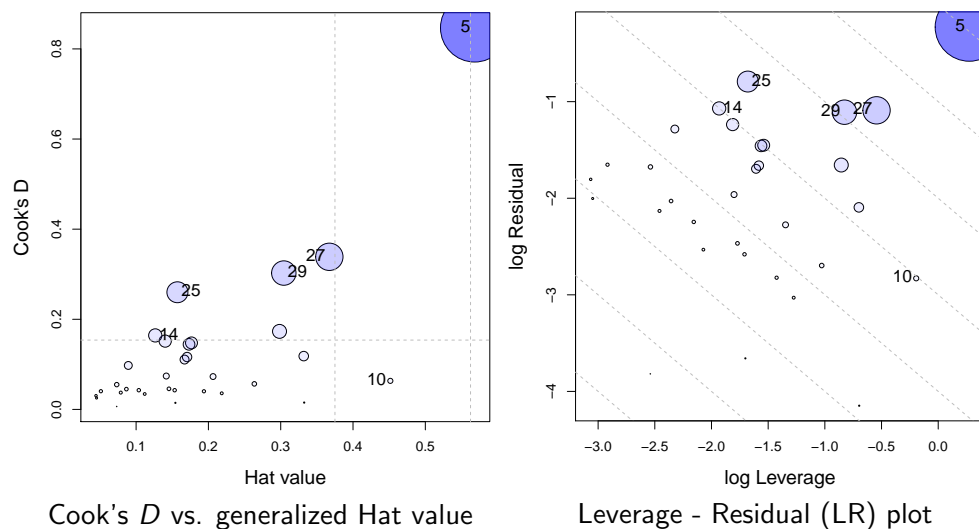
$$H_I = \mathbf{X}_I (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_I^T \quad (2)$$

$$D_I = [\text{vec}(\mathbf{B} - \mathbf{B}_{(I)})]^T [\mathbf{S}^{-1} \otimes (\mathbf{X}^T \mathbf{X})] [\text{vec}(\mathbf{B} - \mathbf{B}_{(I)})] \quad (3)$$

- Provides deletion diagnostics for subsets (I) of size $m \geq 1$.
- e.g., $m = 2$ can reveal cases of **masking** or **joint influence**.
- Extension of `influencePlot()` to the multivariate case.
- A new plot format: **leverage-residual (LR) plots**.

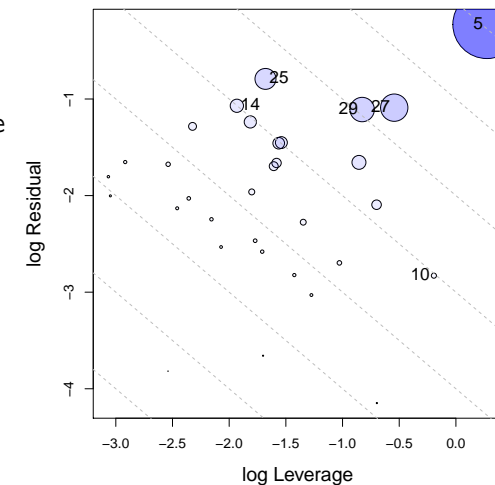
Influence diagnostics for MLMs: Example

For the Rohwer data:



Influence diagnostics for MLMs: LR plots

- Main idea: Influence \sim Leverage (L) \times Residual (R)
- $\mapsto \log(Infl) = \log(L) + \log(R)$
- \mapsto contours of constant influence lie on lines with slope $= -1$.
- Bubble size \sim influence (Cook's D)
- This simplifies interpretation of influence measures



Conclusions: Graphical methods for MLMs

Summary & Opportunities

- **Data ellipse:** visual summary of bivariate relations
 - Useful for multiple-group, MANOVA data
 - Embed in scatterplot matrix: pairwise, bivariate relations
 - Easily extend to show partial relations, robust estimators, etc.
- **HE plots:** visual summary of multivariate tests for MANOVA and MMRA
 - Group means (MANOVA) or 1-df H vectors (MMRA) aid interpretation
 - Embed in HE plot matrix: all pairwise, bivariate relations
 - Extend to show partial relations: HE plot of "adjusted responses"
- **Dimension-reduction techniques:** low-rank (2D) visual summaries
 - Biplot: Observations, group means, biplot data ellipses, variable vectors
 - Canonical HE plots: Similar, but for dimensions of maximal discrimination
- **Beautiful and useful geometries:**
 - Ellipses everywhere; eigenvector-ellipse geometries!
 - Visual representation of significance in MLM
 - Opportunities for other extensions

— FIN —