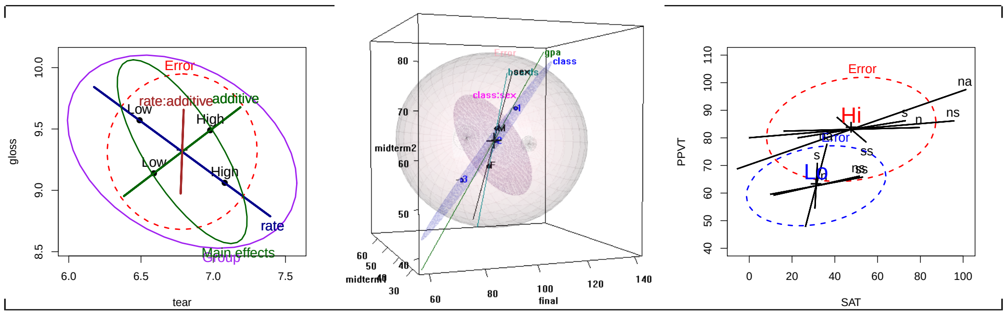
Visualizing multivariate linear models in R

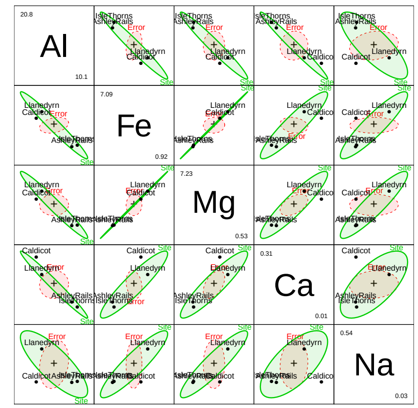
Michael Friendly[1] Matthew Sigal[2]

[1]York University, Toronto [2]Simon Fraser University

useR 2019
Toulouse, July 9–12, 2019



Outline



Slides: `http://datavis.ca/papers/useR2019-2x2.pdf`

Overview: Research topics

Graphical methods for univariate response models well-developed. What about MLMs?

- This talk outlines research on graphical methods for multivariate linear models (MLMs)— extending visualization for multiple regression, ANOVA, and ANCOVA designs to those with several response variables.
- The topics addressed include:
  - Visualizing multivariate tests with Hypothesis–Error (HE) plots in 2D and 3D
  - Low-D views: Generalized canonical discriminant analysis $\rightarrow$ canonical HE plots
  - Visualization methods for tests of equality of covariance matrices in MANOVA designs
  - Extending these methods to robust MLMs
  - Developing multivariate analogs of influence measures and diagnostic plots for MLMs.

Overview: R packages

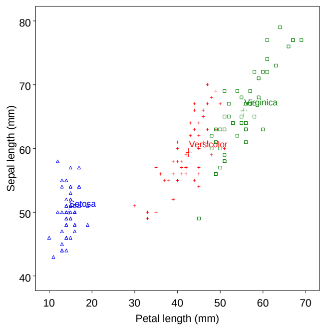
The following R packages implement these methods:

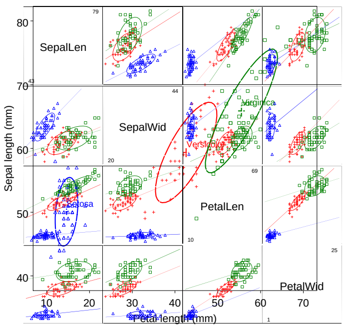
- car package: provides the infrastructure for hypothesis tests (`Anova()`) and tests of linear hypotheses (`linearHypothesis()`) in MLMs, including repeated measures designs.
- heplots package: implements the HE plot framework in 2D (`heplot()`), 3D (`heplot3d()`), and scatterplot matrix form (`pairs.mlm()`). Also provides:
  - `covEllipses()` for covariance ellipses, with optional robust estimation
  - `boxM()` and related methods for testing / visualizing equality of covariance matrices in MANOVA
  - Tutorial vignettes and many data set examples of use
- candisc package: generalized canonical discriminant analysis for an MLM, and associated plot methods.
- mvinfluence package: Multivariate extensions of leverage and influence (Cook's D) and `influencePlot.mlm()` in various forms.
- genridge package: Generalized 2D & 3D ridge regression plots.

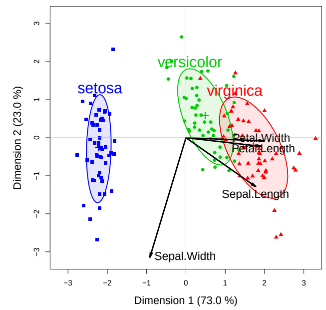
Visual overview: Multivariate data, $\boldsymbol{Y}_{n \times p}$

What we know how to do well (almost)

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



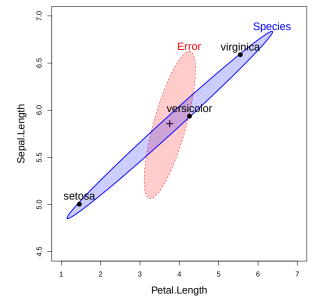


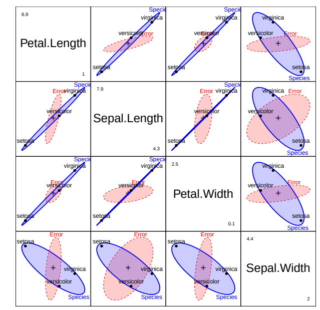


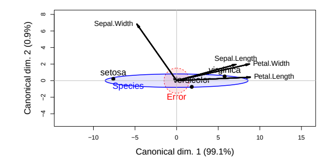
Visual overview: Multivariate linear model, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U}$

What is new here?

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



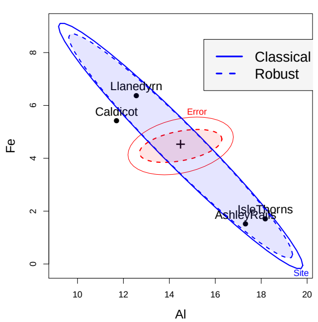


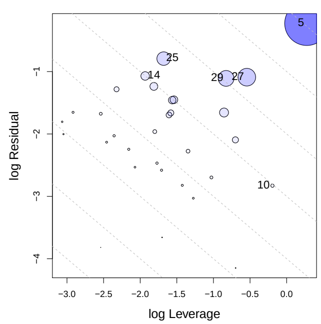


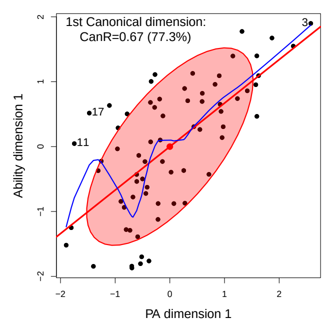
Visual overview: Recent extensions

Extending univariate methods to MLMs:

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



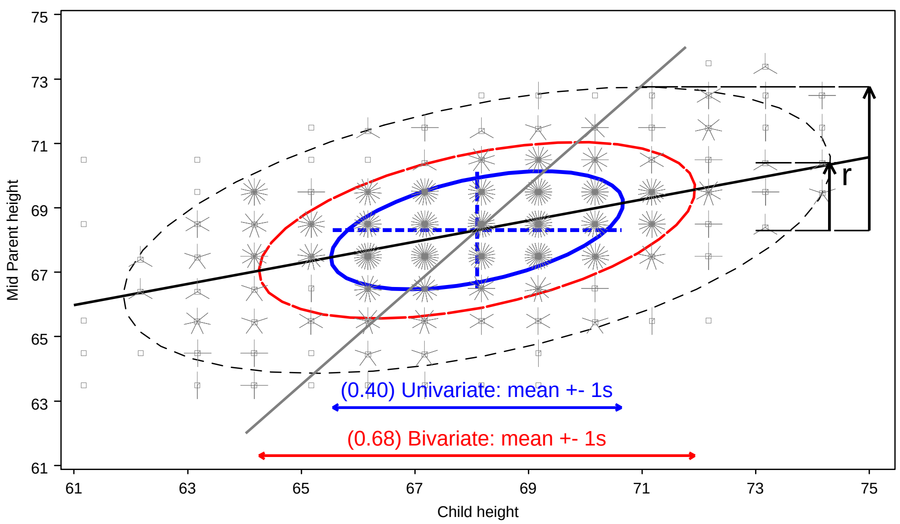




Data ellipsoids: Visually sufficient summaries

- For any p-variable, multivariate normal $\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, the mean vector $\bar{\boldsymbol{y}}$ and sample covariance $\boldsymbol{S}$ are sufficient statistics
- Geometrically, contours of constant density are ellipsoids centered at $\boldsymbol{\mu}$ with size and shape determined by Σ
- $\mapsto$ the data (concentration) ellipsoid, $\mathcal{E}(\bar{\boldsymbol{y}}, \boldsymbol{S})$ is a sufficient visual summary
- Easily robustified by using robust estimators of location and scatter

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 $\leq c^2$,

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

  $\mathbf{S}$ = sample 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: $\bar{y} \pm 1s$
    - $c^2 = \chi^2_2(0.68) = 2.28$: 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, $\mathcal{U} = (\sin \boldsymbol{\theta}, \cos \boldsymbol{\theta})$,

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

  $\mathbf{S}^{1/2}$ = any "square root" of $\mathbf{S}$ (e.g., Cholesky)
- 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**: $\hat{\beta} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1} \mathbf{X}^\mathsf{T}\mathbf{y}$
- **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?

$$F = \frac{SS_H / \mathrm{df}_h}{SS_E / \mathrm{df}_e} = \frac{MS_H}{MS_E} \longrightarrow (MS_H - F \, MS_E) = 0$$

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, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p)$
- **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 $\mathbf{E}$,

$$\mathbf{H} = (\mathbf{C}\widehat{\mathbf{B}})^\mathsf{T} \, [\mathbf{C}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-}\mathbf{C}^\mathsf{T}]^{-1} \, (\mathbf{C}\widehat{\mathbf{B}}) \ ,$$
$$\mathbf{E} = \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{Y}^\mathsf{T}[\mathbf{I} - \mathbf{H}]\mathbf{Y} \ .$$

- Analog of univariate F is

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

- How big is $\mathbf{H}$ relative to $\mathbf{E}$?
  - Latent roots $\lambda_1, \lambda_2, \ldots \lambda_s$ measure the "size" of $\mathbf{H}$ relative to $\mathbf{E}$ in $s = \min(p, df_h)$ 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)
- → One-way MANOVA design, 4 groups, 5 responses

```
R> library(heplots)
R> Pottery
         Site   Al   Fe   Mg   Ca   Na
1    Llanedyrn 14.4 7.00 4.30 0.15 0.51
2    Llanedyrn 13.8 7.08 3.43 0.12 0.17
3    Llanedyrn 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> car::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
```

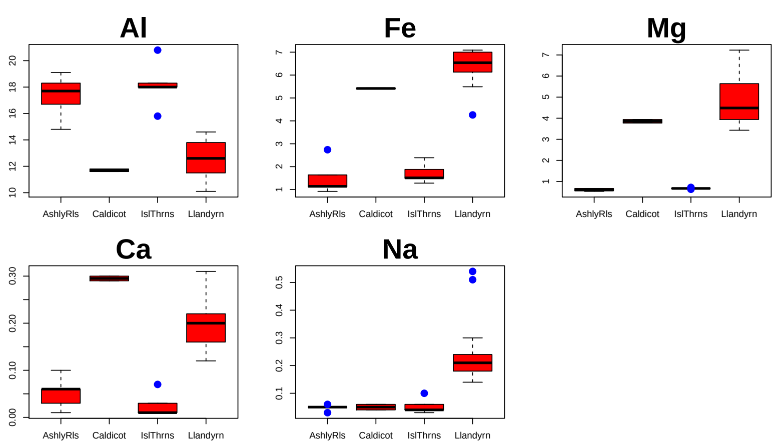
What have we learned?

- **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 response variables to each other
- Do not show *how* variables contribute to multivariate tests

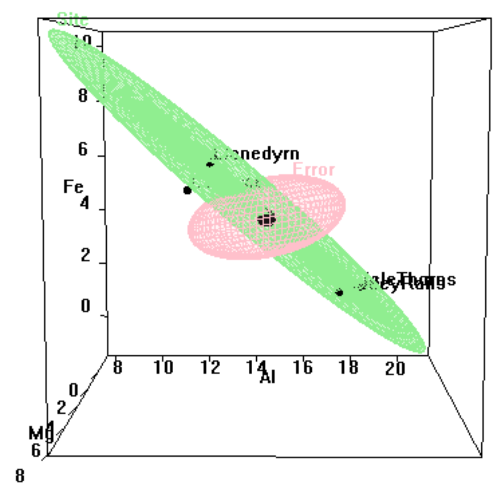


Motivating Example: Romano-British Pottery

Visual answer: HE plot

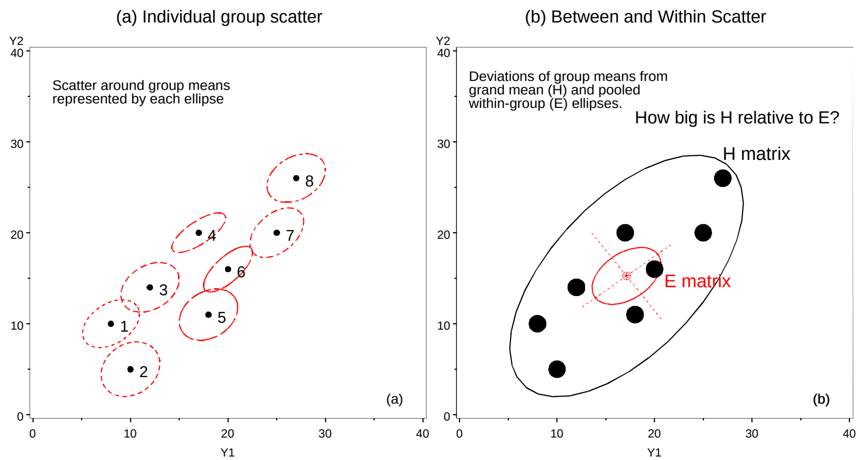
- Shows variation of means (***H***) relative to residual (***E***) variation
- Size and orientation of ***H*** wrt ***E***: *how much* and *how* variables contribute to discrimination
- Evidence scaling: ***H*** is scaled so that it projects outside ***E*** *iff* null hypothesis is rejected.

Run heplot-movie.ppt




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

HE plots: Visualizing *H* and *E* variation



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

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

HE plot details: *H* and *E* matrices

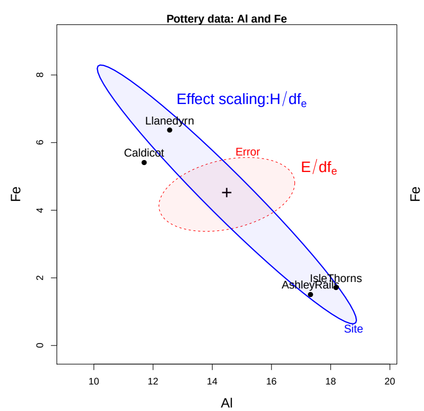
Recall the data on 5 chemical elements in samples of Romano-British pottery from 4 kiln sites:

```
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
```

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

HE plot details: Scaling *H* and *E*

- The E ellipse is divided by $df_e = (n - p) \rightarrow$ data ellipse of residuals
  - Centered at grand means → show factor means in same plot.
- "Effect size" scaling– $\mathbf{H}/df_e \rightarrow$ data ellipse of fitted values.
- "Significance" scaling– H ellipse protrudes beyond 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 ***H*** wrt ***E*** ↦ linear combinations that depart from H_0.



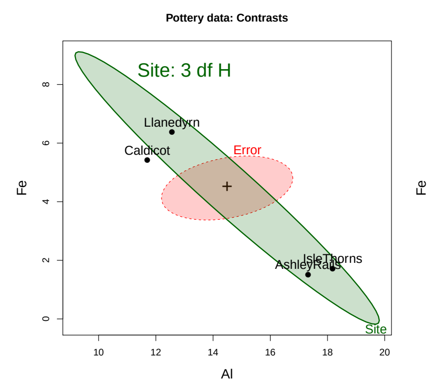

```
R> heplot(pottery.mod, size="effect")
R> heplot(pottery.mod, size="evidence")
```


HE plot details: Contrasts and linear hypotheses

- An overall effect ↦ an ***H*** ellipsoid of $s = \min(p, df_h)$ dimensions
- Linear hypotheses, of rank h, $H_0 : \mathbf{C}_{h \times q} \mathbf{B}_{q \times p} = \mathbf{0}_{h \times p}$ ↦ sub-ellipsoid of dimension h

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

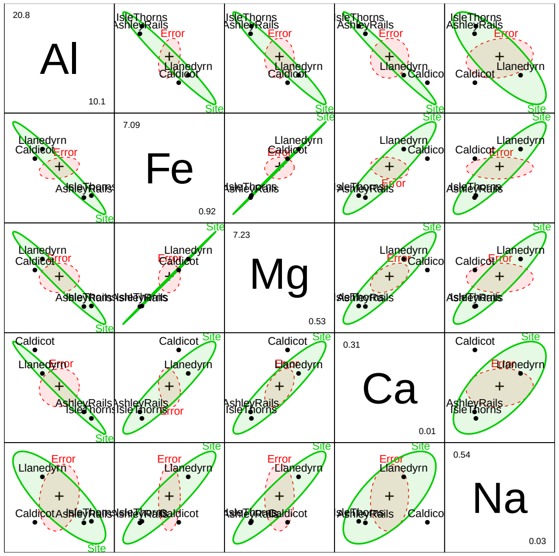
- 1D tests and contrasts ↦ degenerate 1D ellipses (lines)
- Beautiful geometry:
  - Sub-hypotheses are tangent to enclosing hypotheses
  - Orthogonal contrasts form conjugate axes



HE plot matrices: All bivariate views

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

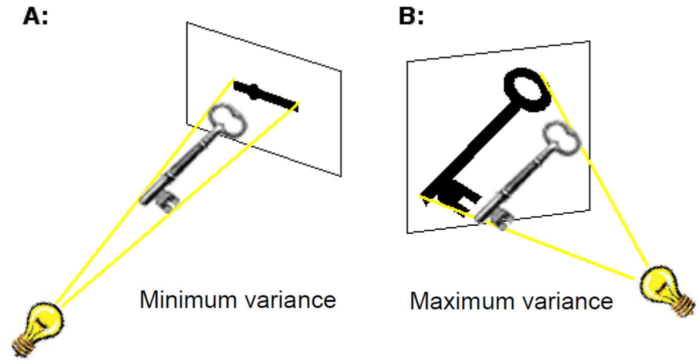
Jump to low-D




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

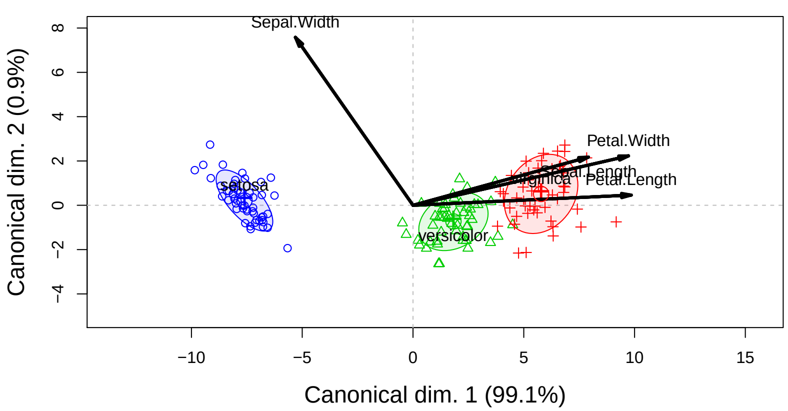

Low-D displays of high-D data

- 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.
- $\rightarrow$ low-D approximation to high-D data



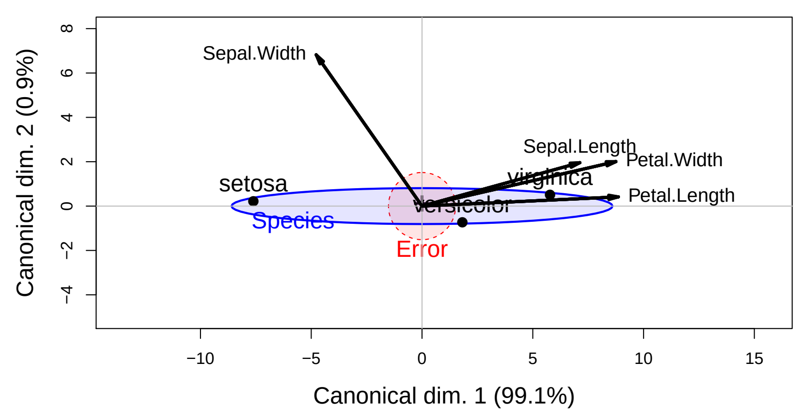
Canonical discriminant HE plots

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



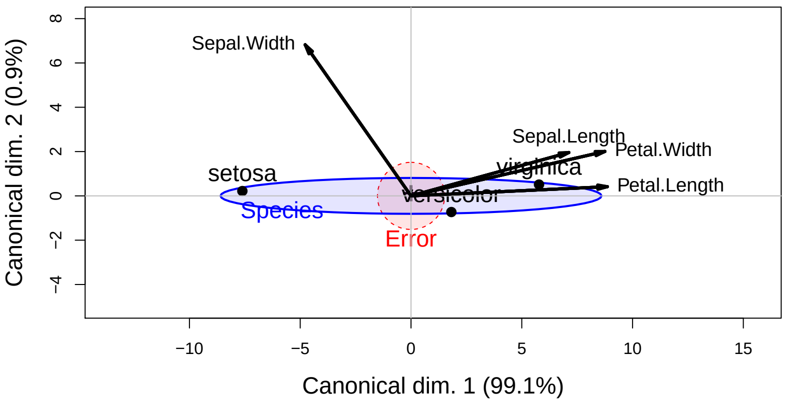
Canonical discriminant HE plots

- Canonical HE plot is just the HE plot of canonical scores, $(\boldsymbol{z}_1, \boldsymbol{z}_2)$ in 2D,
- or, $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3$, in 3D.
- As in biplot, we add vectors to show relations of the $\boldsymbol{y}_i$ response variables to the canonical variates.
- variable vectors here are structure coefficients = correlations of variables with canonical scores.



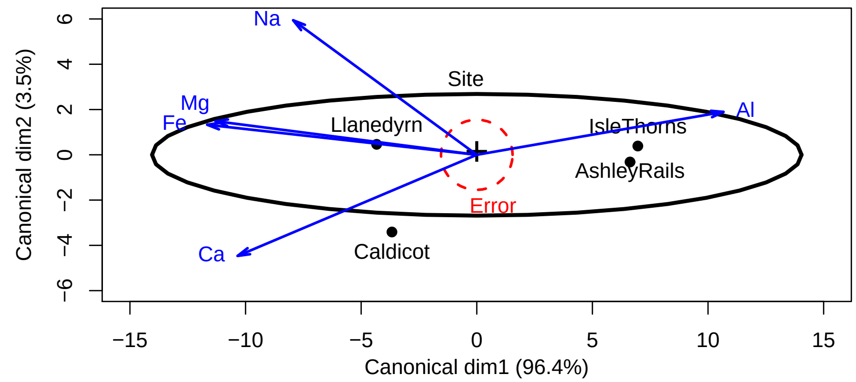
Canonical discriminant HE plots: Properties

- 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 distingiushing (Caldicot, Llandryn) from (Isle Thorns, Ashley Rails): 96.4% of mean variation
- Na and Ca contribute an additional 3.5%. End of story!



Run heplot-movie.ppt

Robust MLMs

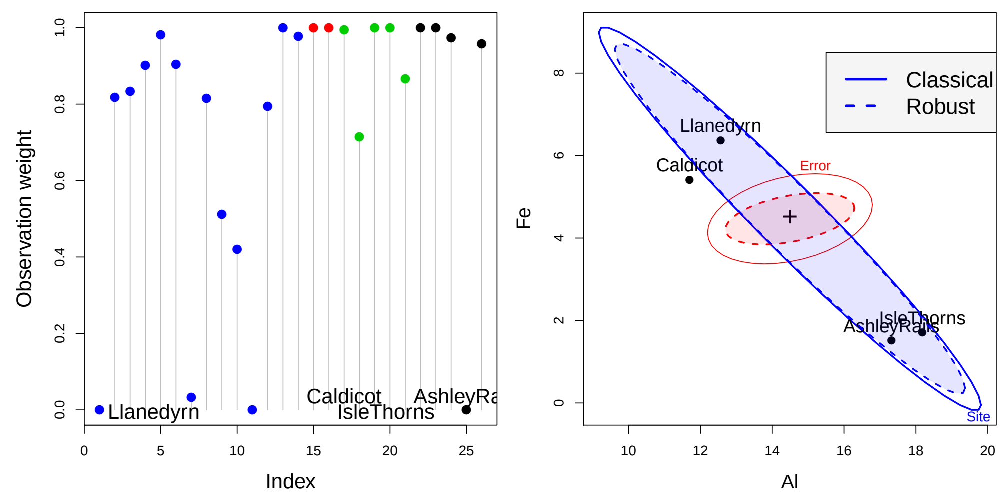
- R has a large collection of packages dealing with robust estimation:
  - `robust::lmrob()`, `MASS::rlm()`, for *univariate* LMs
  - `robust::glmrob()` for univariate *generalized* LMs
  - High breakdown-bound methods for robust *PCA* and robust covariance estimation
  - However, none of these handle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

$$D^2 = (\boldsymbol{Y} - \widehat{\boldsymbol{Y}})^{\mathsf{T}} \boldsymbol{S}_{\text{trob}}^{-1} (\boldsymbol{Y} - \widehat{\boldsymbol{Y}}) \sim \chi^2_p \quad (1)$$

  - This fully extends the `"mlm"` class
  - Compatible with other `mlm` extensions: `car::Anova()` and `heplot()`.

Robust MLMs: Example

For the Pottery data:



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

Influence diagnostics for MLMs

- Influence measures & diagnostic plots well-developed for *univariate* LMs
  - Influence measures: Cook's D, DFFITS, dfbetas, etc.
  - Diagnostic plots: Index plots, `car::influencePlot()` for LMs
  - However, these are 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:

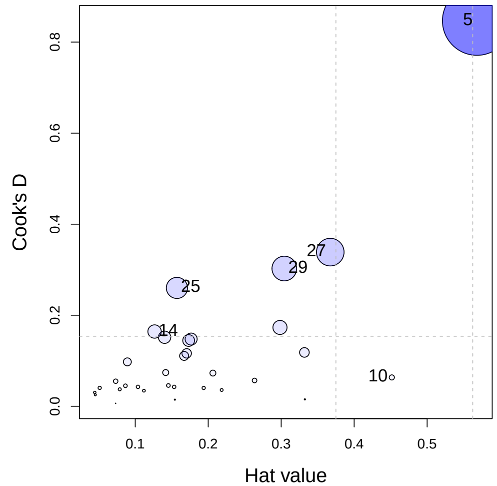
$$H_I = \boldsymbol{X}_I(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}_I^\mathsf{T} \qquad (2)$$

$$D_I = [vec(\boldsymbol{B} - \boldsymbol{B}_{(I)})]^\mathsf{T}[\boldsymbol{S}^{-1} \otimes (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})][vec(\boldsymbol{B} - \boldsymbol{B}_{(I)})] \qquad (3)$$

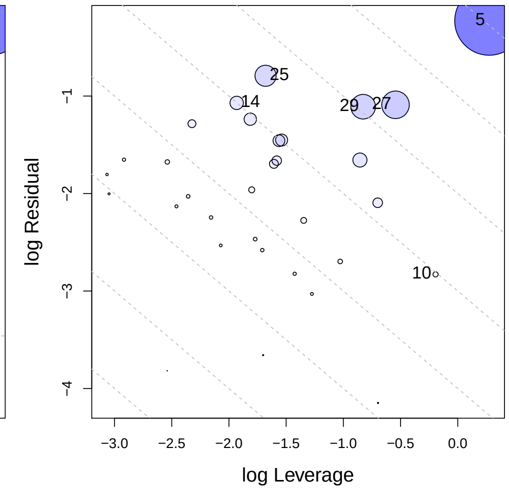
  - Provides deletion diagnostics for *subsets* (I) of size $m \geq 1$.
  - e.g., $m = 2$ can reveal cases of masking or joint influence.
  - Extension of `influencePlot()` to the multivariate case.
  - A new plot format: leverage-residual (LR) plots (McCulloch & Meeter, 1983)

Influence diagnostics for MLMs: Example

For the Rohwer data:



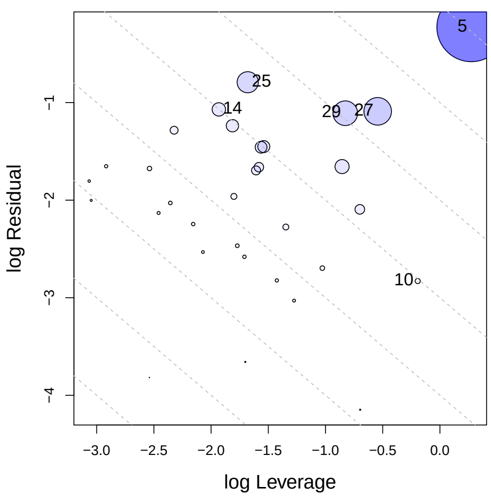
Cook's D vs. generalized Hat value



Leverage - Residual (LR) plot

Influence diagnostics for MLMs: LR plots

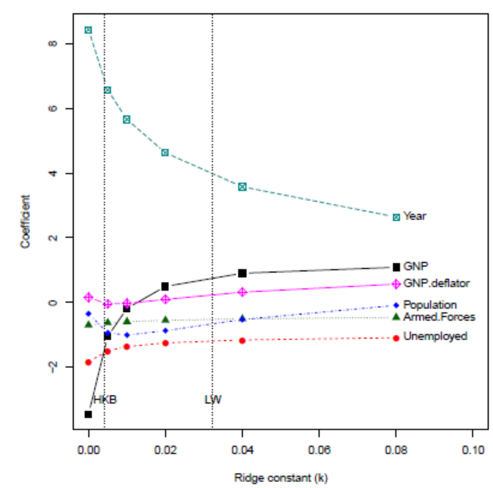
- Main idea: Influence ∼ Leverage (L) × Residual (R)
- $\mapsto \log(Infl) = \log(L) + \log(R)$
- $\mapsto$ contours of constant influence lie on lines with slope = -1.
- Bubble size ∼ influence (Cook's D)
- This simplifies interpretation of influence measures



Ridge regression plots

Shrinkage methods often use ridge trace plots to visualize effects

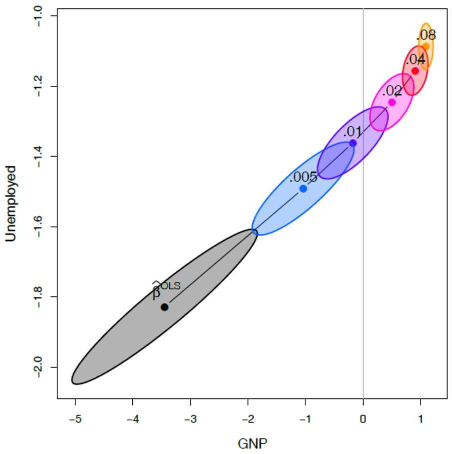
- Typical: univariate line plot of β_k vs. shrinkage, k
- What can you see here regarding bias vs. precision?
- This is the wrong graphic form, for a multivariate problem!
- Goal: visualize $\widehat{\beta}_k$ vs. $\widehat{\text{Var}}(\widehat{\beta}_k)$

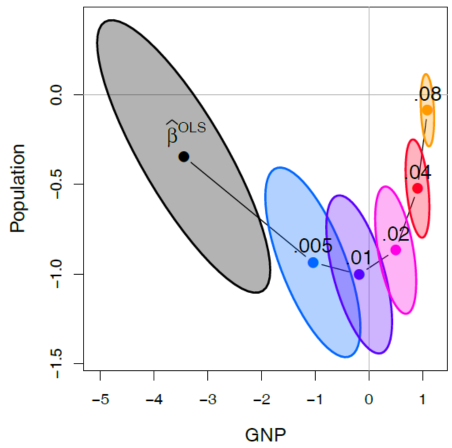


Generalized ridge trace plots

Rather than plotting just the univariate trajectories of β_k vs. K, plot the 2D (3D) confidence ellipsoids over the same range of k.

- Centers of the ellipsoids are $\widehat{\beta_k}$ – same info as in univariate plot.
- Can see how change in one coefficient is related to changes in others.
- Relative size & shape of ellipsoids show directly effect on precision.





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 et Merci —