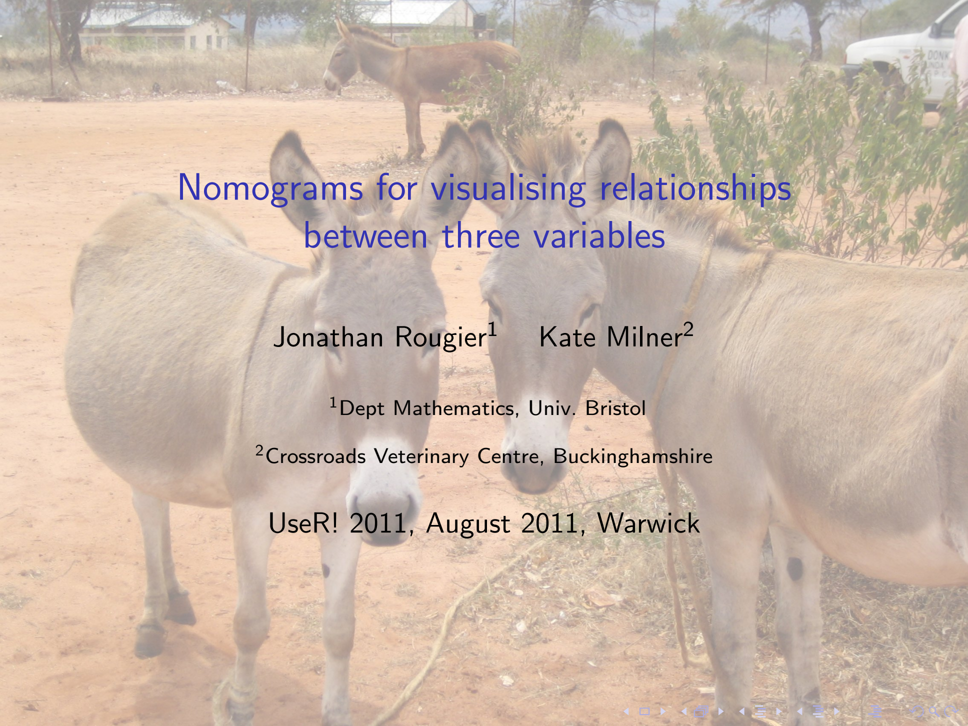


Nomograms for visualising relationships between three variables

Jonathan Rougier[1]   Kate Milner[2]

[1]Dept Mathematics, Univ. Bristol

[2]Crossroads Veterinary Centre, Buckinghamshire

UseR! 2011, August 2011, Warwick

Background

A donkey drawn by my housemate Caroline (in the pub).

Background

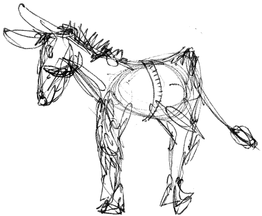
A donkey drawn by my housemate Caroline (in the pub).

This donkey is not enjoying being weighed.

Background

A donkey drawn by my housemate Caroline (in the pub).

This donkey is not enjoying being weighed.



A happy baby donkey being measured.

Usual practice

The standard practice is to fit a relationship

$$\log(\texttt{Weight}) = a + b\log(\texttt{HeartGirth}) + c\log(\texttt{Height})$$

to adult donkeys in good condition, and possibly other relationships for juveniles and donkeys in poor condition. What value can we statisticians add?

Usual practice

The standard practice is to fit a relationship

$$\log(\texttt{Weight}) = a + b\log(\texttt{HeartGirth}) + c\log(\texttt{Height})$$

to adult donkeys in good condition, and possibly other relationships for juveniles and donkeys in poor condition. What value can we statisticians add?

1. Explicit inclusion of factors for `Age`, `Gender`, and `BCS` (Body Condition Score);

Usual practice

The standard practice is to fit a relationship

$$\log(\texttt{Weight}) = a + b\log(\texttt{HeartGirth}) + c\log(\texttt{Height})$$

to adult donkeys in good condition, and possibly other relationships for juveniles and donkeys in poor condition. What value can we statisticians add?

1. Explicit inclusion of factors for `Age`, `Gender`, and `BCS` (Body Condition Score);
2. Box-Cox assessment of the appropriate transformation of the lefthand side (`boxcox` in the MASS package);

Usual practice

The standard practice is to fit a relationship

$$\log(\texttt{Weight}) = a + b\log(\texttt{HeartGirth}) + c\log(\texttt{Height})$$

to adult donkeys in good condition, and possibly other relationships for juveniles and donkeys in poor condition. What value can we statisticians add?

1. Explicit inclusion of factors for `Age`, `Gender`, and `BCS` (Body Condition Score);
2. Box-Cox assessment of the appropriate transformation of the lefthand side (`boxcox` in the MASS package);
3. Initial model to include interactions, then stepwise reduction to maximise AIC (`stepAIC` in the MASS package).

Building the statistical model

Box-Cox plot for transformations of the response favours square root

Building the statistical model

Backwards stepwise deletion removes all interaction terms :) and Gender completely

```
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
sqrt(Weight) ~ BCSis + Gender + Age + log(HeartGirth) + log(Height) +
    log(HeartGirth):log(Height) + BCSis:log(HeartGirth) + Gender:log(HeartGirth) +
    Age:log(HeartGirth) + BCSis:log(Height) + Gender:log(Height) +
    Age:log(Height)

Final Model:
sqrt(Weight) ~ BCSis + Age + log(HeartGirth) + log(Height)


                           Step Df   Deviance Resid. Df Resid. Dev        AIC
1                                                   504   78.14041  -972.7873
2          - Age:log(HeartGirth)  5 0.37630656       509   78.51672  -980.1883
3        - BCSis:log(HeartGirth)  4 0.49082973       513   79.00755  -984.8168
4            - BCSis:log(Height)  4 0.41453445       517   79.42208  -989.9858
5              - Age:log(Height)  5 0.91895494       522   80.34104  -993.7620
6           - Gender:log(Height)  2 0.13986420       524   80.48090  -996.8210
7 - log(HeartGirth):log(Height)  1 0.00927524       525   80.49018  -998.7587
8       - Gender:log(HeartGirth)  2 0.31844543       527   80.80862 -1000.6226
9                       - Gender  2 0.06633122       529   80.87496 -1004.1787
```

Building the statistical model

Resulting model has additive adjustments for BCS and Age

```
Call:
lm(formula = sqrt(Weight) ~ BCSis + Ageis + log(HeartGirth) +
    log(Height), data = donk, subset = subset)

Residuals:
      Min        1Q    Median        3Q       Max
-1.016797 -0.275575 -0.005298  0.255089  1.519246

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -58.89411    2.42162 -24.320  < 2e-16 ***
BCSis1.5         -0.49820    0.17939  -2.777  0.00568 **
BCSis2           -0.24978    0.08253  -3.026  0.00260 **
BCSis3.5          0.37485    0.05833   6.426 2.91e-10 ***
BCSis4            0.57031    0.11024   5.173 3.27e-07 ***
Ageis<2yo        -0.35353    0.07676  -4.605 5.16e-06 ***
Ageis5-10yo       0.19782    0.06255   3.162  0.00165 **
Ageis>10yo        0.27681    0.05070   5.459 7.35e-08 ***
log(HeartGirth)  10.22732    0.50604  20.211  < 2e-16 ***
log(Height)       4.84926    0.60029   8.078 4.45e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.392 on 531 degrees of freedom
Multiple R-squared: 0.8724, Adjusted R-squared: 0.8703
F-statistic: 403.5 on 9 and 531 DF,  p-value: < 2.2e-16
```

Nomogram for our donkeys

Our statistical estimate of `Weight` is

$$\texttt{Weight} = \Big(-58.9^{\dagger} + 10.2 \log \texttt{HeartGirth} + 4.8 \log \texttt{Height} \Big)^2$$

where † indicates adjustments to be made for `BCS` and `Age`. *How do we turn this into something that can be used in the field?*

- Most statisticians would immediately think of a contour plot, which would work for any relationship of the form $f(u, v) = w$. This requires two straight lines and an interpolation.

Nomogram for our donkeys

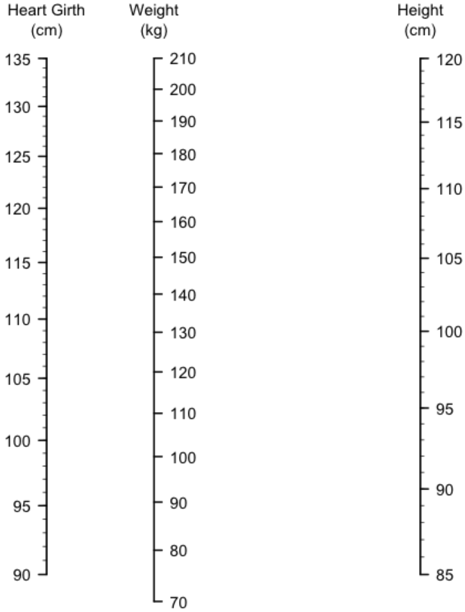
Our statistical estimate of `Weight` is

$$\texttt{Weight} = \Big(-58.9^{\dagger} + 10.2 \log \texttt{HeartGirth} + 4.8 \log \texttt{Height} \Big)^2$$

where † indicates adjustments to be made for `BCS` and `Age`. *How do we turn this into something that can be used in the field?*

- Most statisticians would immediately think of a contour plot, which would work for any relationship of the form $f(u, v) = w$. This requires two straight lines and an interpolation.
- For a large subset of such relationships, though, we can construct a nomogram, which needs one straight line and no interpolation.

Nomogram for our donkeys

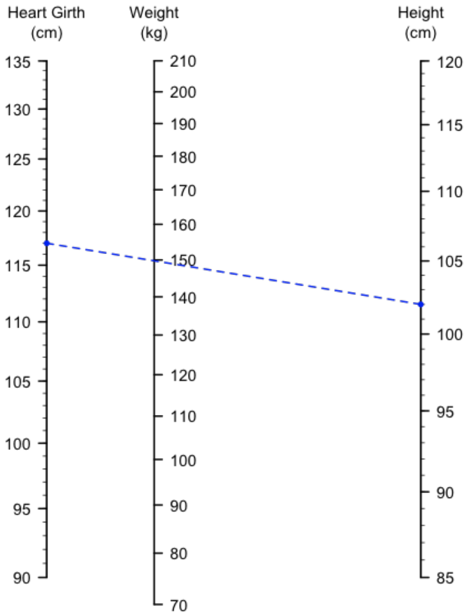


Additive corrections:

```
BCS: 1.5,    -11kg
     2,       -6kg
     3.5,    +10kg
     4,      +16kg

Age: <2yo,     -7kg
     5-10yo, +5kg
     >10yo,   +7kg
```

Nomogram for our donkeys



Additive corrections:

```
BCS: 1.5,   -11kg
     2,      -6kg
     3.5,   +10kg
     4,     +16kg

Age: <2yo,    -7kg
     5-10yo, +5kg
     >10yo,  +7kg
```

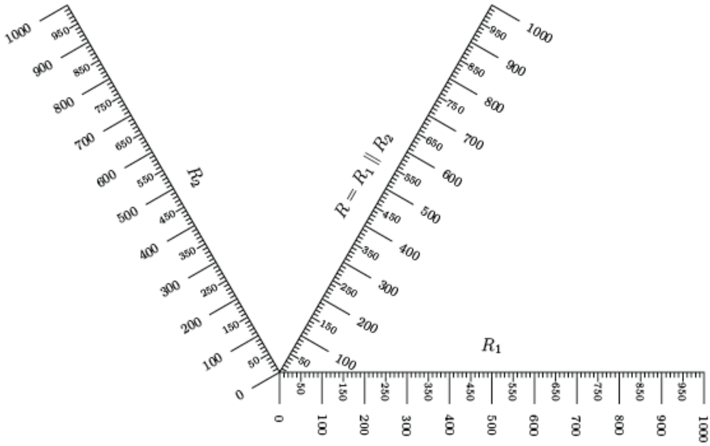
A healthy (BCS 2.5 or 3) 2-5yo donkey with a `HeartGirth` of 117cm and a `Height` of 102cm has a predicted weight of about 150kg.

Digression on nomograms

Nomograms are visual tools for representing the relationship between three or more variables, in such a way that the value of one variable can be inferred from the values of the others by drawing a straight line.

- $f_1(u) + f_2(v) = f_3(w)$ gives a parallel scale-nomogram, like ours;
- We could also have used an N chart, used for $f_1(u)/f_2(v) = f_3(w)$;
- Proportional nomograms can handle more than three variables, e.g. in two stages using a pivot;
- An entire theory based around determinants allows the construction of nomograms for much more general relationships; typically these are curved scale nomograms.

Digression on nomograms

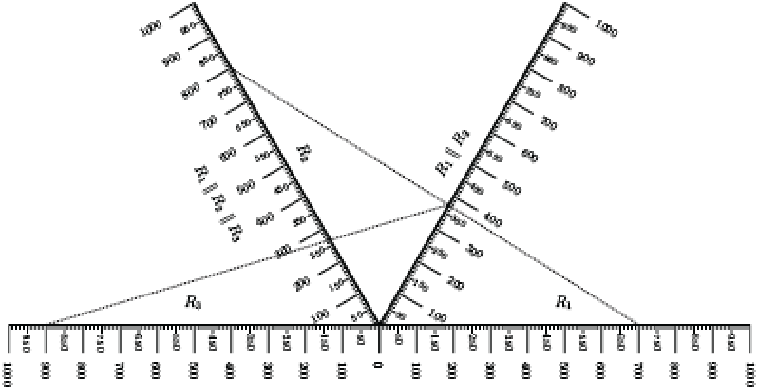


Equivalent Resistance of Two Resistors in Parallel

$$1/R = 1/R_1 + 1/R_2$$

All figures from Ron Doerfler, 2009, The Lost Art of Nomography, *The UMAP Journal*, **30**(4), pp. 457-493.

Digression on nomograms

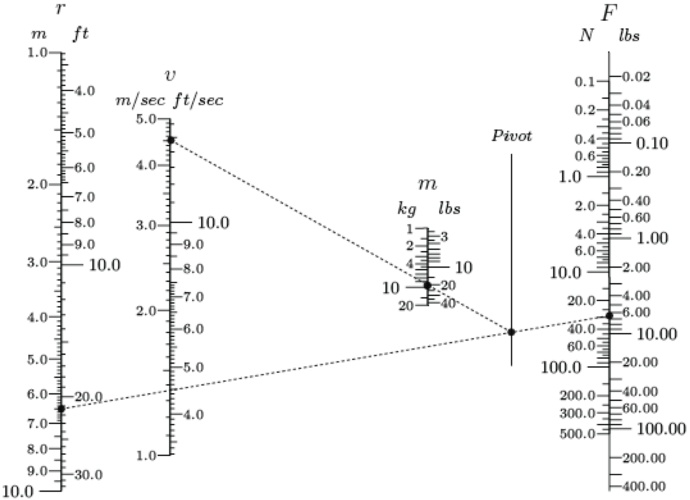


Equivalent Resistance of Three Resistors in Parallel

$$1/R = 1/R_1 + 1/R_2 + 1/R_3$$

All figures from Ron Doerfler, 2009, The Lost Art of Nomography, *The UMAP Journal*, **30**(4), pp. 457-493.

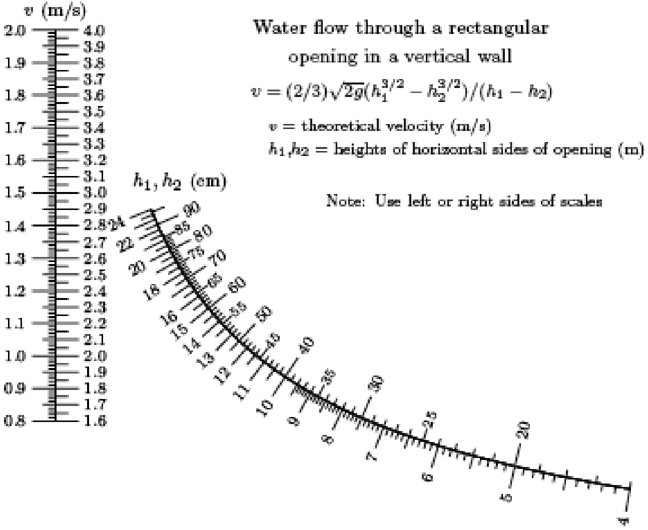
Digression on nomograms



Centripetal Force: $F = mv^2/r$

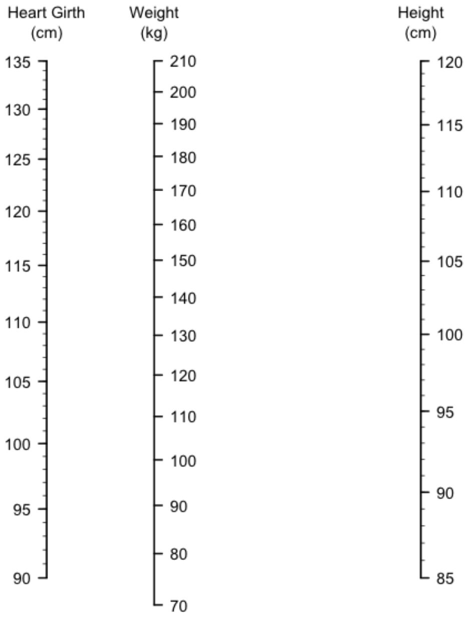
All figures from Ron Doerfler, 2009, The Lost Art of Nomography, *The UMAP Journal*, **30**(4), pp. 457-493.

Digression on nomograms



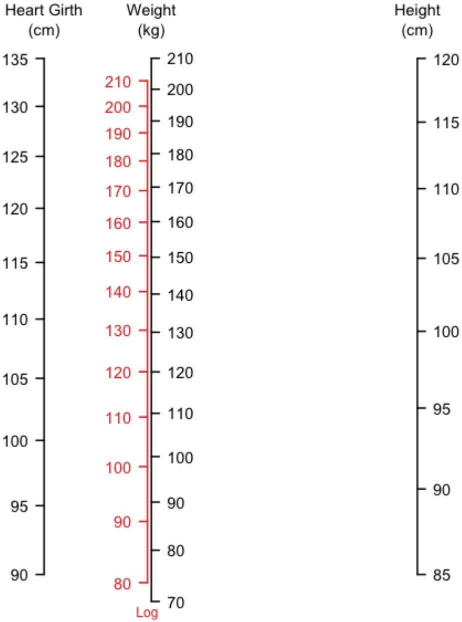
All figures from Ron Doerfler, 2009, The Lost Art of Nomography, *The UMAP Journal*, **30**(4), pp. 457-493.

Back to the donkeys!



What is the effect of replacing sqrt(`Weight`) with log(`Weight`), which would be the more usual transformation?

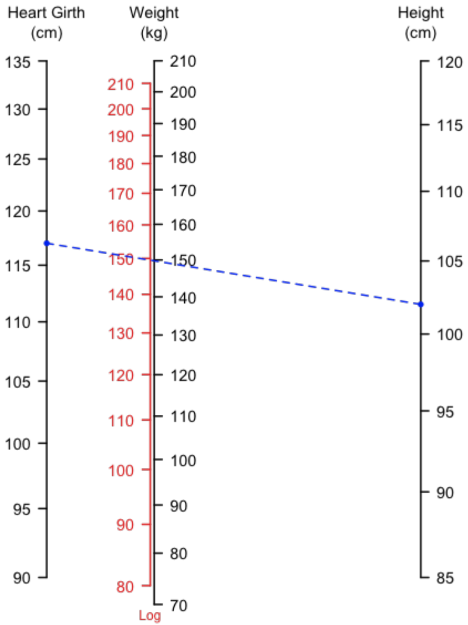
Back to the donkeys!



What is the effect of replacing sqrt(`Weight`) with log(`Weight`), which would be the more usual transformation?

Gives slightly higher weights (∼5kg) for small and large donkeys. This difference is smaller than the residual standard deviation, which is 10kg.

Back to the donkeys!

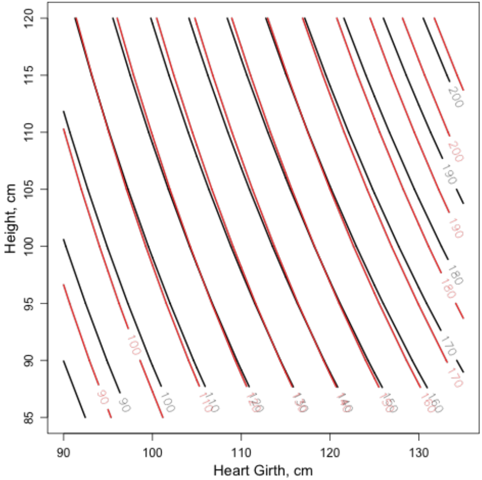


What is the effect of replacing sqrt(`Weight`) with log(`Weight`), which would be the more usual transformation?

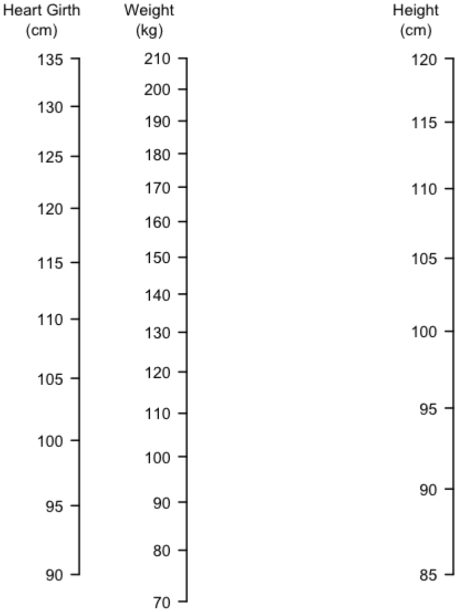
Gives slightly higher weights (∼5kg) for small and large donkeys. This difference is smaller than the residual standard deviation, which is 10kg.

Back to the donkeys!

Things are a lot less clear if we try to visualise this using a contour plot.

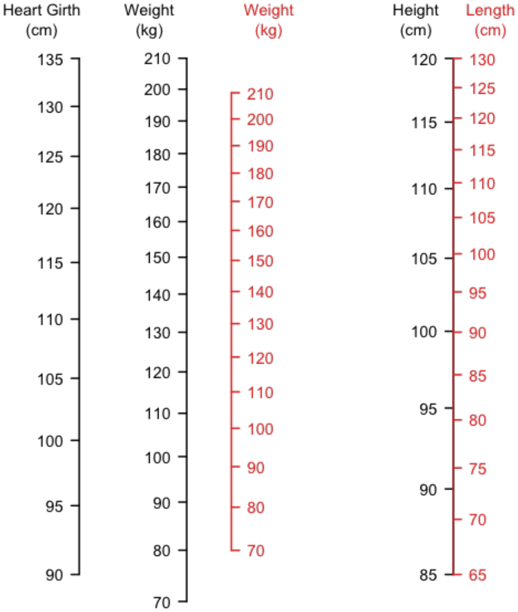


Different relationships on one plot



`Height` and `Length` seem to be interchangeable; so could estimate `Weight` with either.

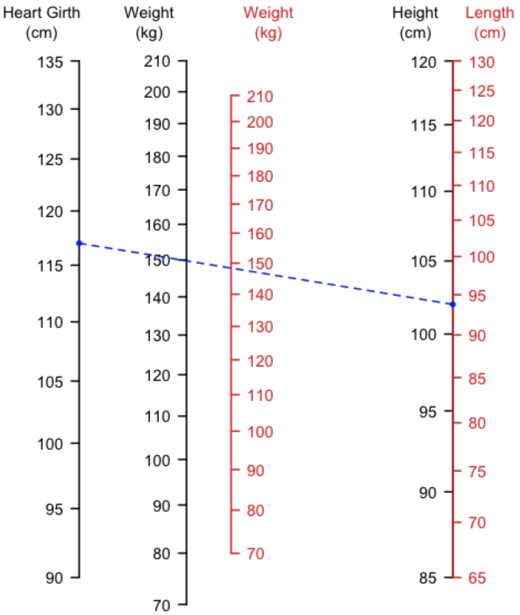
Different relationships on one plot



Height and Length seem to be interchangeable; so could estimate Weight with either.

Estimate using Length can be added to existing nomogram, to give vets the choice of which measurement to make.

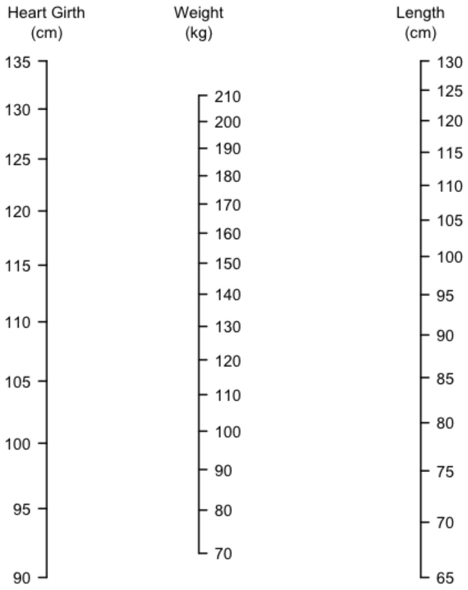
Different relationships on one plot



`Height` and `Length` seem to be interchangeable; so could estimate `Weight` with either.

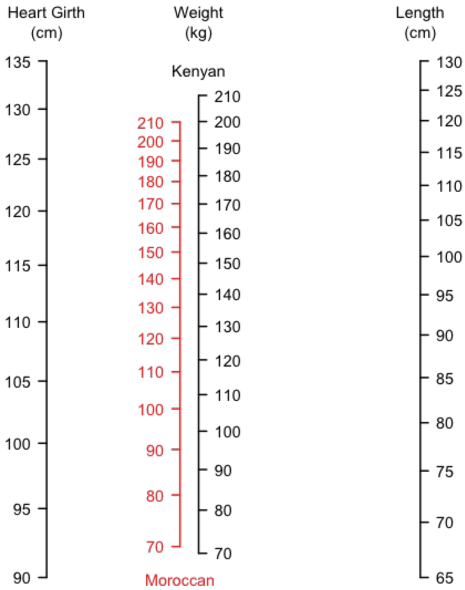
Estimate using `Length` can be added to existing nomogram, to give vets the choice of which measurement to make.

Different types of donkey



Different types of donkey can be displayed on the same plot. Here are our Kenyan donkeys, shown with a `Length` covariate.

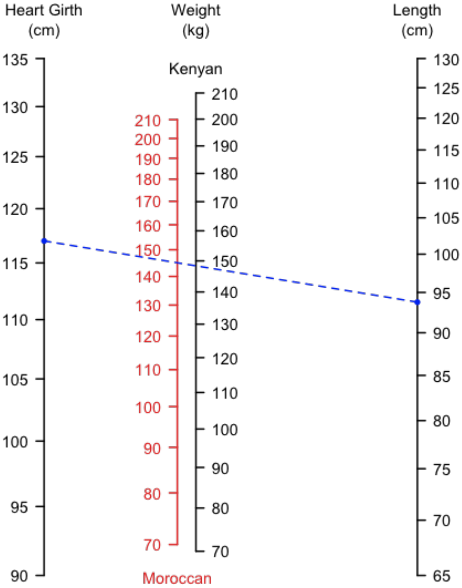
Different types of donkey



Different types of donkey can be displayed on the same plot. Here are our Kenyan donkeys, shown with a `Length` covariate.

This is for Moroccan donkeys. They tend to be a bit lighter for the same size.

Different types of donkey



Different types of donkey can be displayed on the same plot. Here are our Kenyan donkeys, shown with a `Length` covariate.

This is for Moroccan donkeys. They tend to be a bit lighter for the same size.

Summary

Visualisation is an important part of both data analysis and statistical communication.

- For relating three variables, contour plots will always work, but where they are available, nomograms might be clearer and simpler to use.
- Our donkey nomogram will be used by practicing vets in Kenya, but it has also been a useful tool for us in model choice and model comparison.
- Nomograms are also available for some relationships between four or more variables.
- One catch: Contour plots can be overlaid on a field showing predictive uncertainties. Unfortunately it is not as easy to visualise predictive uncertainty with a nomogram.

Resources

Ron Doerfler, 2009, The Lost Art of Nomography, *The UMAP Journal*, **30**(4), pp. 457-493.
http://myreckonings.com/wordpress/wp-content/uploads/JournalArticle/The_Lost_Art_of_Nomography.pdf

Ron Doerfler, Creating Nomograms with the *PyNomo* Software, Version 1.1 for PyNomo Release 0.2.2.
http://www.myreckonings.com/pynomo/CreatingNomogramsWithPynomo.pdf

Leif Roschier, 2009, http://www.pynomo.org/