Initial EDA - 136 farms

A T Chamberlain

2024-07-24

# NOTES

# 24 July - increased from 105 to 136 herds - these are all routine HPRs from Emma

# 25 July - start sensible outlier checking - have removed 6+9 for BF + culling = 5%

# 13 Aug - 149 files, better data processing - hopefully finalversion of data

## [1] "C:/ATC\_data/HPRs/Paper\_Eric\_OEE+LCA"

## Rows: 149 Columns: 53  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Cull\_L6, Cull\_L6over  
## dbl (51): farmID, annualMY, herdSize, annualMYFPCM, AAFC, CI, culling, avLac...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## [1] "Herd number"

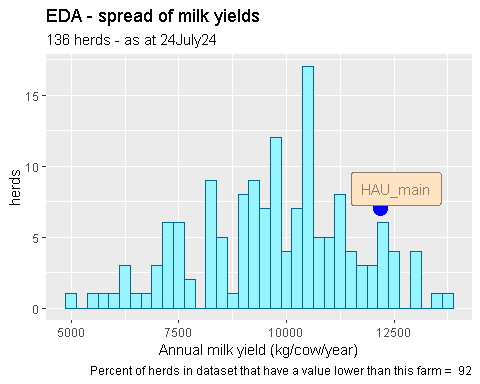
## [1] 1

## [1] "Herd Name"

## [1] "HAU\_main"

## [1] "Herd ID"

## [1] 1



## [1] "Value for this herd :"

## [1] 12182

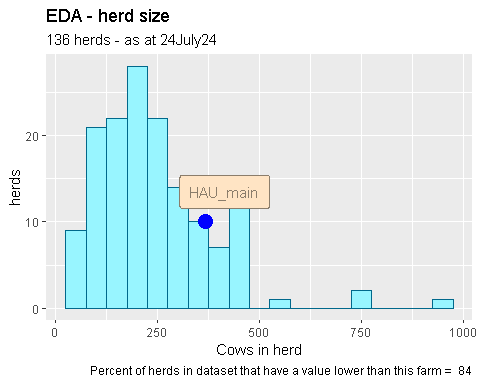
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5057 8467 9811 9739 10911 13653

## [1] 5057

## [1] 13653

## [1] 38

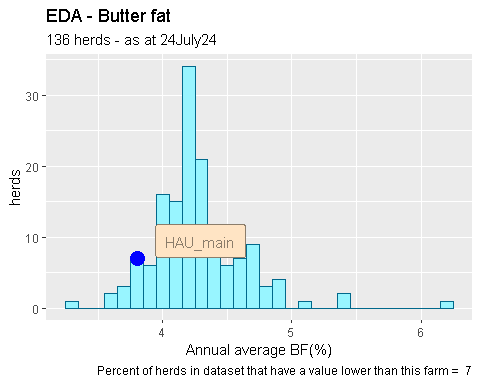
## [1] 79



## [1] "Value for this herd :"

## [1] 367

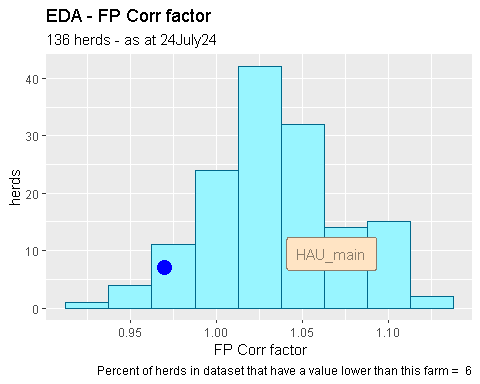
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 32 140 220 240 294 944



## [1] "Value for this herd :"

## [1] 3.8007

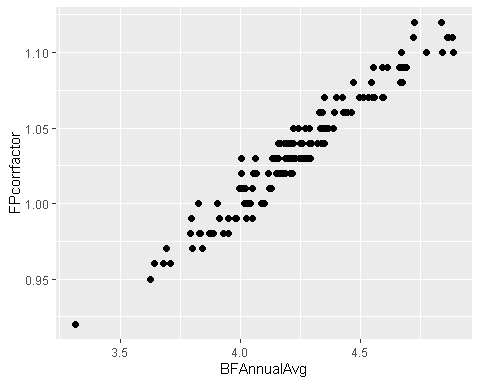
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3.314 4.056 4.216 4.267 4.401 6.211



## [1] "Value for this herd :"

## [1] 0.97

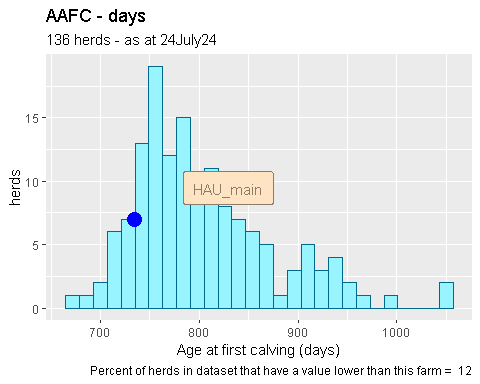
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.920 1.010 1.030 1.034 1.060 1.120



## [1] 0.9686058

Have rejected 4 herds with butter fats over 5%. Previously, when selecting herds for this database dropped herds known to be NOT Hol/Fr from herd name, location (Channel Islands), personal knowledge.

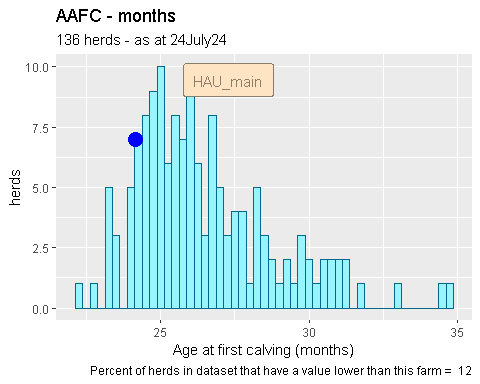
# look at distribution of the main driving KPI’s - excl MY - is any trimming needed?



## [1] "Value for this herd :"

## [1] 734

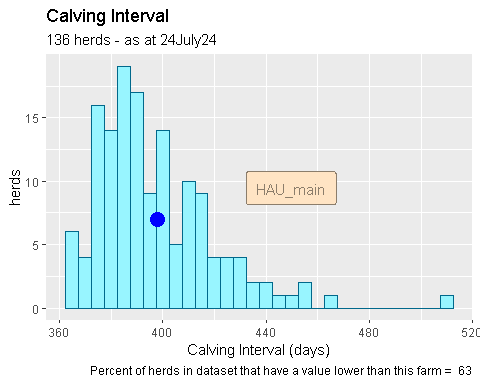
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 673.0 752.0 788.0 804.8 844.0 1055.0



## [1] "Value for this herd :"

## [1] 24.14474

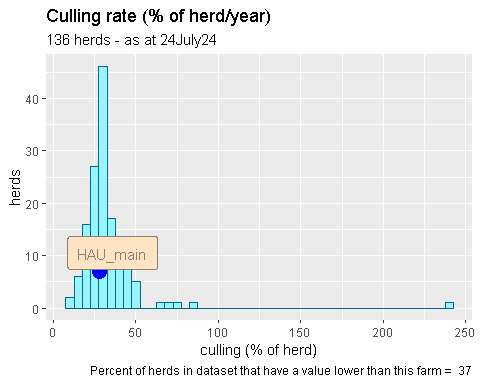
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 22.14 24.74 25.92 26.47 27.76 34.70



## [1] "Value for this herd :"

## [1] 398

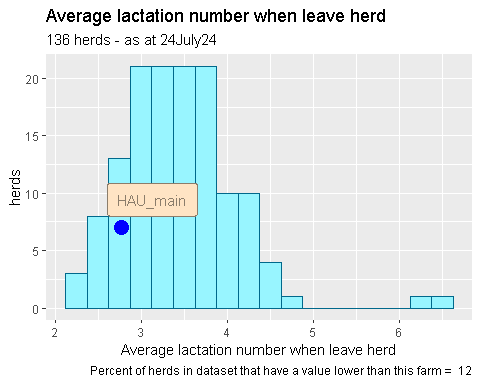
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 364 381 392 397 408 511



## [1] "Value for this herd :"

## [1] 27.7

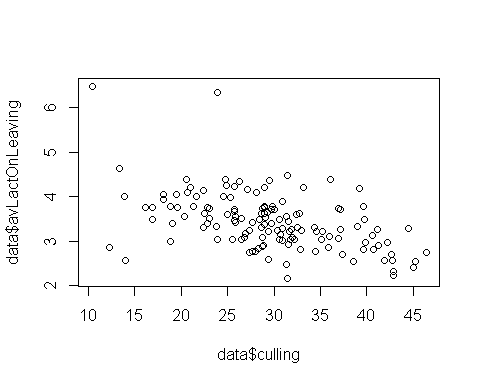
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10.40 25.50 29.50 32.91 36.10 240.90



## [1] "Value for this herd :"

## [1] 2.77

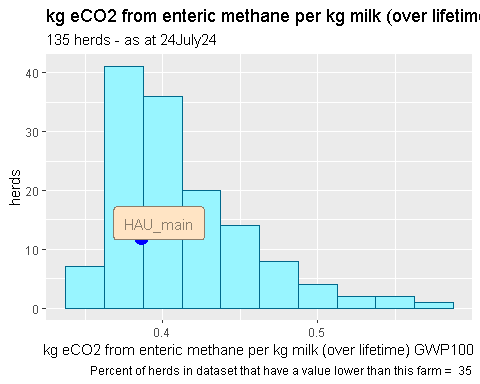
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.170 3.035 3.400 3.442 3.755 6.470



## [1] -0.4916912

Ten herds rejected as culling rates over 50%. Initially 149, 4 lost for high butterfat, 10 lost for high culling. 14 removed out of 149 = 9.4% rejected.

# finally look at distribution of eCO2 / kg FP Corr milk



## [1] "Value for this herd :"

## [1] 0.386

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.3448 0.3798 0.4018 0.4126 0.4337 0.5751

SUMMARY - have deleted 6 herds as they looked like Channel Island breed and 9 herds as culling rate was too high due to TB culls, selling down-calved heifers, or downsizing / closing up / retiring.

Find out who are highest and lowest emitters

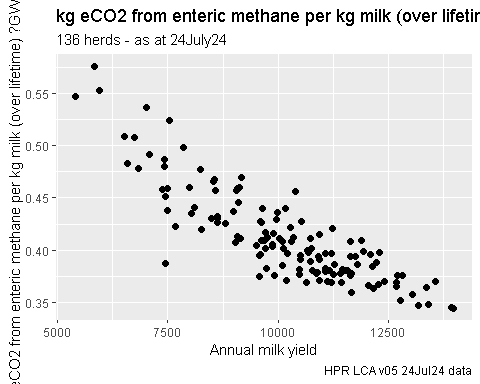
## [1] 0.3448

## [1] 0.5751

## [1] 30

## [1] 102

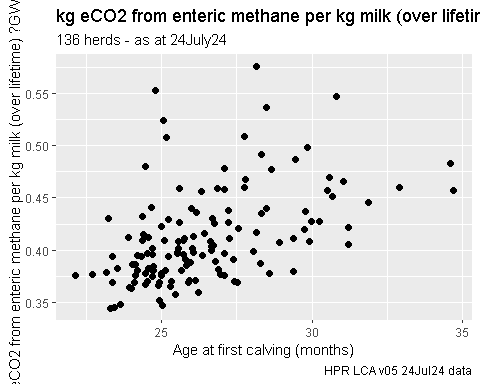
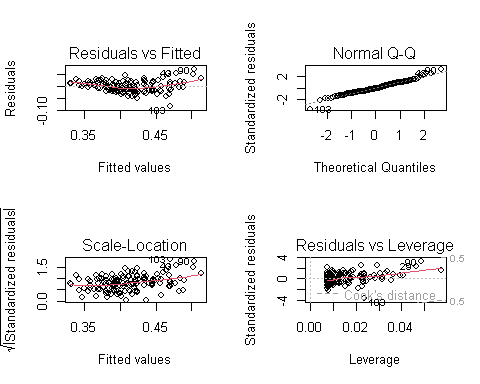
Finally - in EDA - look at some simple regressions



## [1] -0.860406

##   
## Call:  
## lm(formula = eCO2GWP100Total ~ annualMYFPCM, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.082295 -0.014592 -0.002353 0.015282 0.071151   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.277e-01 1.123e-02 55.90 <2e-16 \*\*\*  
## annualMYFPCM -2.124e-05 1.091e-06 -19.47 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02311 on 133 degrees of freedom  
## Multiple R-squared: 0.7403, Adjusted R-squared: 0.7383   
## F-statistic: 379.1 on 1 and 133 DF, p-value: < 2.2e-16

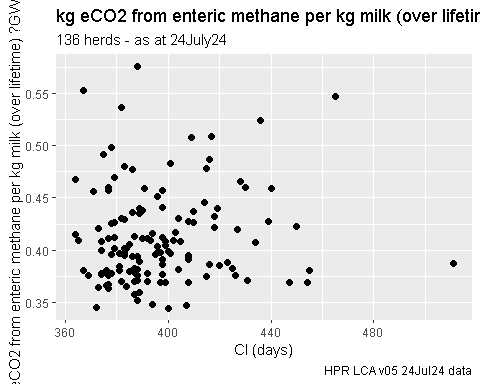
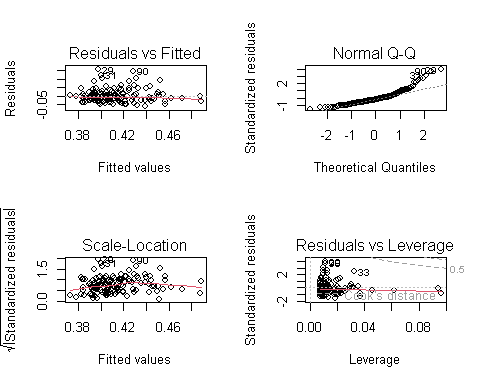
## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.628 0.0112 55.9 2.96e-94  
## 2 annualMYFPCM -0.0000212 0.00000109 -19.5 9.25e-41



## [1] 0.4866639

##   
## Call:  
## lm(formula = eCO2GWP100Total ~ AAFCMonths, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.060181 -0.024991 -0.007394 0.013307 0.154959   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.169394 0.038001 4.458 1.74e-05 \*\*\*  
## AAFCMonths 0.009194 0.001431 6.425 2.17e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.03961 on 133 degrees of freedom  
## Multiple R-squared: 0.2368, Adjusted R-squared: 0.2311   
## F-statistic: 41.28 on 1 and 133 DF, p-value: 2.168e-09

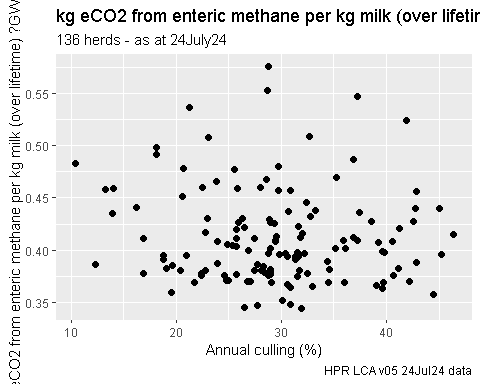
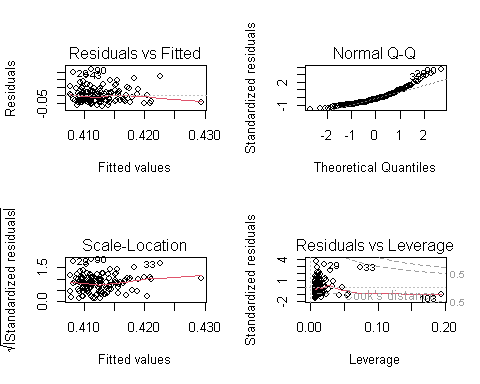
## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.169 0.0380 4.46 0.0000174   
## 2 AAFCMonths 0.00919 0.00143 6.42 0.00000000217



## [1] 0.07458322

##   
## Call:  
## lm(formula = eCO2GWP100Total ~ CI, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.068237 -0.032218 -0.009248 0.021102 0.163830   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3541418 0.0678305 5.221 6.68e-07 \*\*\*  
## CI 0.0001472 0.0001707 0.863 0.39   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04521 on 133 degrees of freedom  
## Multiple R-squared: 0.005563, Adjusted R-squared: -0.001914   
## F-statistic: 0.744 on 1 and 133 DF, p-value: 0.3899

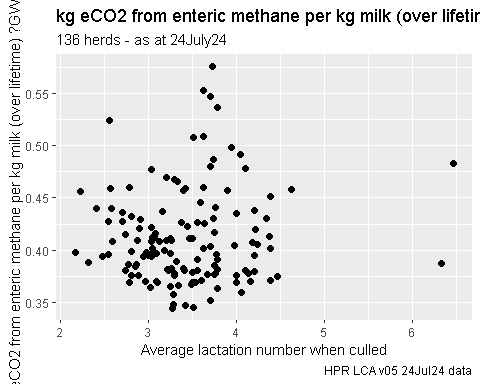
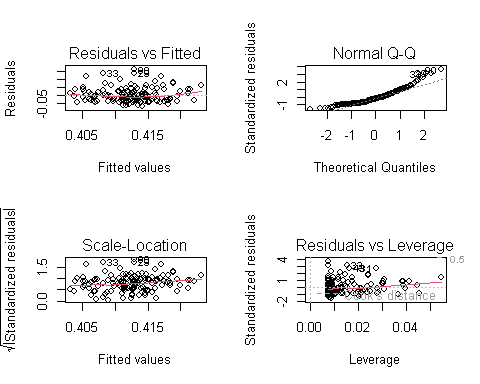
## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.354 0.0678 5.22 0.000000668  
## 2 CI 0.000147 0.000171 0.863 0.390



## [1] -0.09329401

##   
## Call:  
## lm(formula = eCO2GWP100Total ~ culling, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.069190 -0.034711 -0.008615 0.021301 0.162185   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.4288845 0.0156054 27.483 <2e-16 \*\*\*  
## culling -0.0005545 0.0005131 -1.081 0.282   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04514 on 133 degrees of freedom  
## Multiple R-squared: 0.008704, Adjusted R-squared: 0.00125   
## F-statistic: 1.168 on 1 and 133 DF, p-value: 0.2818

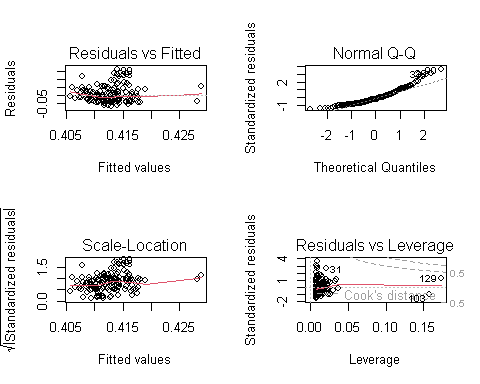
## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.429 0.0156 27.5 1.07e-56  
## 2 culling -0.000554 0.000513 -1.08 2.82e- 1



## [1] 0.07688382

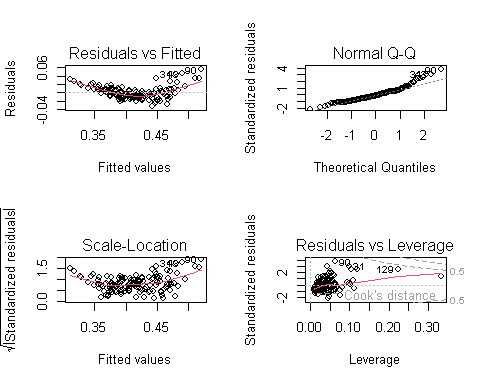
##   
## Call:  
## lm(formula = eCO2GWP100Total ~ +avLactOnLeaving, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.06792 -0.03373 -0.01094 0.02051 0.16099   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.393941 0.021287 18.507 <2e-16 \*\*\*  
## avLactOnLeaving 0.005407 0.006080 0.889 0.375   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04521 on 133 degrees of freedom  
## Multiple R-squared: 0.005911, Adjusted R-squared: -0.001563   
## F-statistic: 0.7909 on 1 and 133 DF, p-value: 0.3754

## # A tibble: 2 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.394 0.0213 18.5 1.30e-38  
## 2 avLactOnLeaving 0.00541 0.00608 0.889 3.75e- 1



##   
## Call:  
## lm(formula = eCO2GWP100Total ~ annualMYFPCM + AAFCMonths + CI +   
## avLactOnLeaving + culling, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.031530 -0.010077 -0.002357 0.007614 0.055051   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.029e-01 3.347e-02 23.991 < 2e-16 \*\*\*  
## annualMYFPCM -2.448e-05 8.886e-07 -27.549 < 2e-16 \*\*\*  
## AAFCMonths 3.199e-03 6.298e-04 5.080 1.29e-06 \*\*\*  
## CI -3.990e-04 6.113e-05 -6.528 1.39e-09 \*\*\*  
## avLactOnLeaving -2.280e-02 2.518e-03 -9.057 1.81e-15 \*\*\*  
## culling 3.330e-04 2.020e-04 1.649 0.102   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01524 on 129 degrees of freedom  
## Multiple R-squared: 0.8904, Adjusted R-squared: 0.8861   
## F-statistic: 209.5 on 5 and 129 DF, p-value: < 2.2e-16

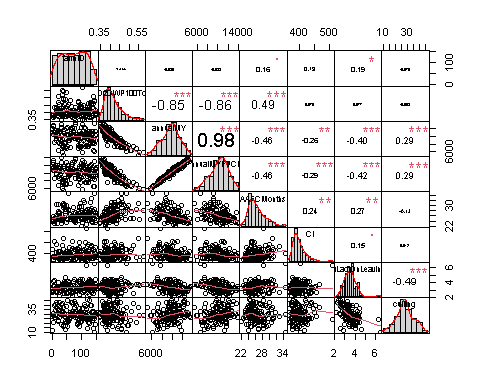
## # A tibble: 6 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 0.803 0.0335 24.0 2.15e-49  
## 2 annualMYFPCM -0.0000245 0.000000889 -27.5 6.93e-56  
## 3 AAFCMonths 0.00320 0.000630 5.08 1.29e- 6  
## 4 CI -0.000399 0.0000611 -6.53 1.39e- 9  
## 5 avLactOnLeaving -0.0228 0.00252 -9.06 1.81e-15  
## 6 culling 0.000333 0.000202 1.65 1.02e- 1



## farmID eCO2GWP100Total annualMY annualMYFPCM AAFCMonths CI  
## farmID 1.00 -0.03 -0.09 -0.08 0.16 0.13  
## eCO2GWP100Total -0.03 1.00 -0.85 -0.86 0.49 0.07  
## annualMY -0.09 -0.85 1.00 0.98 -0.46 -0.26  
## annualMYFPCM -0.08 -0.86 0.98 1.00 -0.46 -0.29  
## AAFCMonths 0.16 0.49 -0.46 -0.46 1.00 0.24  
## CI 0.13 0.07 -0.26 -0.29 0.24 1.00  
## avLactOnLeaving 0.19 0.08 -0.40 -0.42 0.27 0.15  
## culling -0.08 -0.09 0.29 0.29 -0.13 0.01  
## avLactOnLeaving culling  
## farmID 0.19 -0.08  
## eCO2GWP100Total 0.08 -0.09  
## annualMY -0.40 0.29  
## annualMYFPCM -0.42 0.29  
## AAFCMonths 0.27 -0.13  
## CI 0.15 0.01  
## avLactOnLeaving 1.00 -0.49  
## culling -0.49 1.00

## farmID eCO2GWP100Total annualMY annualMYFPCM AAFCMonths CI  
## farmID 1.00 -0.03 -0.09 -0.08 0.16 0.13  
## eCO2GWP100Total -0.03 1.00 -0.85 -0.86 0.49 0.07  
## annualMY -0.09 -0.85 1.00 0.98 -0.46 -0.26  
## annualMYFPCM -0.08 -0.86 0.98 1.00 -0.46 -0.29  
## AAFCMonths 0.16 0.49 -0.46 -0.46 1.00 0.24  
## CI 0.13 0.07 -0.26 -0.29 0.24 1.00  
## avLactOnLeaving 0.19 0.08 -0.40 -0.42 0.27 0.15  
## culling -0.08 -0.09 0.29 0.29 -0.13 0.01  
## avLactOnLeaving culling  
## farmID 0.19 -0.08  
## eCO2GWP100Total 0.08 -0.09  
## annualMY -0.40 0.29  
## annualMYFPCM -0.42 0.29  
## AAFCMonths 0.27 -0.13  
## CI 0.15 0.01  
## avLactOnLeaving 1.00 -0.49  
## culling -0.49 1.00  
##   
## n= 135   
##   
##   
## P  
## farmID eCO2GWP100Total annualMY annualMYFPCM AAFCMonths CI   
## farmID 0.6883 0.3203 0.3407 0.0665 0.1261  
## eCO2GWP100Total 0.6883 0.0000 0.0000 0.0000 0.3899  
## annualMY 0.3203 0.0000 0.0000 0.0000 0.0022  
## annualMYFPCM 0.3407 0.0000 0.0000 0.0000 0.0007  
## AAFCMonths 0.0665 0.0000 0.0000 0.0000 0.0043  
## CI 0.1261 0.3899 0.0022 0.0007 0.0043   
## avLactOnLeaving 0.0300 0.3754 0.0000 0.0000 0.0018 0.0740  
## culling 0.3860 0.2818 0.0007 0.0007 0.1418 0.8891  
## avLactOnLeaving culling  
## farmID 0.0300 0.3860   
## eCO2GWP100Total 0.3754 0.2818   
## annualMY 0.0000 0.0007   
## annualMYFPCM 0.0000 0.0007   
## AAFCMonths 0.0018 0.1418   
## CI 0.0740 0.8891   
## avLactOnLeaving 0.0000   
## culling 0.0000

## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter  
  
## Warning in par(usr): argument 1 does not name a graphical parameter

 A very simple 5 component multiple regression model had a adj R2 of 89%, F=209 on df(5,129), p<0.001. Only culling was not significant in the model - I assume avLactOnLeaving took out most of this variables effect.

This is not the same as sensitivity analysis which I would do with the model I have

. Find an ‘average herd’ or typical herd - how? .. normalise each variable and express as SND’s then look for the herd with the combined / average min absolute SND.

. set all variables into model at the average values . vary one at a time to the decile values

identify several ‘typical’ herds and so sens analysis on each herd and look at difference in results.

This has been done in another R file.