Student Datasets

CM515 2025 Students

2025-01-26

Dataset previews

Amanda

The dataset I am using for the course is a subset of saliva samples obtained from dairy calves to assess cortisone and cortisol concentrations (ng/mL) under various environmental conditions. I obtained a random subset from the larger dataset and would like to look at the relationship between cortisone and cortisol concentrations.

```
amanda_tsv = read.table('student_dataset_previews/amanda.txt', header=T)
head(amanda_tsv)
```

```
##                                                    ID cortisone cortisol
## 1 100_81723-Observation-1-07/29/2024-Bovine-Saliva_231      4.17     2.34
## 2 101_81726-Observation-1-07/17/2024-Bovine-Saliva_245      0.73     0.26
## 3 103_81706-Observation-1-07/29/2024-Bovine-Saliva_108      0.81     0.28
## 4 105_81723-Observation-1-07/25/2024-Bovine-Saliva_174      0.85     0.34
## 5 106_81696-Observation-2-07/25/2024-Bovine-Saliva_223      0.20     0.07
## 6 107_81711-Observation-2-07/25/2024-Bovine-Saliva_123      0.62     0.23
```

Anais

RNA-seq of A.thaliana lines with a gene silenced related to herbicide metabolism (dicamba). The genes are IAA and PIF4.

Andrea

CFU readings I measured over time.

```
library(readxl)
andrea_Dataset <- read_excel("assignment_submissions/submissions2/andrea_Dataset_CFU_1st_experiment.xls
head(andrea_Dataset)
```

```
## # A tibble: 6 x 3
##        CFU  HOUR replicate
##      <dbl> <dbl>     <dbl>
## 1  9700000     0         1
## 2  8500000     0         2
## 3 11200000     0         3
## 4  7200000    24         1
## 5  6400000    24         2
## 6  8700000    24         3
```

Ben

It is a mRNA riboseq data set that Naly would like processed.

Danielle

The data set I am using represents the area of detection of different recombinant PIN proteins expressed in Brachypodium.

```
danielle = read_excel('student_dataset_previews/danielle.xlsx')
head(danielle)
```

```
## # A tibble: 6 x 8
##   Slice  Count `Total Area` `Average Size` `%Area`  Mean `measured values` ratio
##   <chr>  <dbl>        <dbl>          <dbl>   <dbl> <dbl>              <dbl> <dbl>
## 1 2016-~     1     2151704.       2151704.    22.7   255          2118464. 0.985
## 2 2016-~     1     2697147        2697147     28.4   255          2686583. 0.996
## 3 2016-~     1     1904559.       1904559.    20.1   255          1895881. 0.995
## 4 2016-~     1     2028616        2028616     21.4   255          2019356. 0.995
## 5 2016-~     1     2335320.       2335320.    24.6   255          2291438. 0.981
## 6 2016-~     2     2273409.       1136704.    24.0   255          2165911. 0.953
```

Gianna

Cut&Run differential bind output for Rosenberg lab (Aedees mosquito sugar fed v bloodfed day 3)

```
gianna_data = read.table("student_dataset_previews/gianna.csv", header=T, sep=",")
head(gianna_data)
```

```
##        chr     start       end width     Conc Conc_SF_K9Me3 Conc_BF_K9Me3
## 1 AaegL5_3 258080712 258087502  6791 3.373904      1.831654      4.102173
## 2 AaegL5_1 101754512 101825290 70779 6.817945      7.401226      5.822978
## 3 AaegL5_1  86485705  86498529 12825 4.047927      2.670812      4.739473
## 4 AaegL5_1  90639907  90643546  3640 1.413452      2.227344      0.000000
## 5 AaegL5_1  18717177  18772401 55225 6.309490      6.725228      5.723128
## 6 AaegL5_1  91819581  91827670  8090 3.860381      4.404741      2.975754
##        Fold      p.value          FDR
## 1 -2.270519 3.964220e-08 6.747102e-05
## 2  1.578248 2.142869e-07 1.823580e-04
## 3 -2.068661 7.688135e-07 4.361740e-04
## 4  2.227344 4.410870e-06 1.876825e-03
## 5  1.002100 9.244169e-06 3.146715e-03
## 6  1.428987 1.267626e-05 3.595832e-03
```

Shea

TB activation and incident untargeted metabolomics datasets

```
shea_data = read.table('student_dataset_previews/shea.txt', header=T)
head(shea_data)
```

```
##              Sample CSU.TB.02194 CSU.TB.02195 CSU.TB.02196 CSU.TB.02197
## 1             Label   Activation   Activation   Activation   Activation
## 2  100.07552_5.2746  0.988878646  0.929592479  1.230075823  0.846958704
## 3 100.07555_10.0655  1.072707754  0.972337959  1.145245399   0.99710456
## 4  100.07559_9.0135  1.276779295  1.240561696  1.040565234  1.183827088
## 5  102.09116_2.2705  0.860385746  0.608960043   0.92641621   0.21655187
## 6  103.03881_2.2947  0.993136722  0.896292734  1.046614642  0.900744849
##   CSU.TB.02198 CSU.TB.02199 CSU.TB.02200 CSU.TB.02201 CSU.TB.02202 CSU.TB.02203
## 1   Activation   Activation   Activation   Activation   Activation   Activation
## 2  1.058476834  1.062198022  0.964872475  1.040683587  0.963484266  1.109678119
```

```
## 3  1.067688043  1.039705129  0.981718602  1.118647885  0.994805557  1.111371123
## 4  0.998019901  1.183751183   1.07106745  1.135168974  1.033876146  1.070423337
## 5  0.657987944  0.853232713  0.889497225  0.934174229  0.743258731  1.016095315
## 6  1.075314974  0.990186768  0.859808729  1.017314476  0.923077296  1.012097003
##   CSU.TB.02204 CSU.TB.02205 CSU.TB.02206 CSU.TB.02207 CSU.TB.02423 CSU.TB.02424
## 1   Activation   Activation   Activation   Activation     Incident     Incident
## 2  1.040747992  0.901633501  0.785603227  1.040975737  1.008980578  1.246783421
## 3  1.012514793  1.023301641  0.854504481  1.108249051  1.098254182  1.250355756
## 4  0.961533243  1.111873091  0.740912643  1.156160931  1.286495664  1.245246004
## 5  0.922961145  0.891997426  0.833197134  0.994492078  0.974429321  1.088515534
## 6    0.9974415  1.004305917  0.958854668  1.013699432  1.001160937  1.074107294
##   CSU.TB.02425 CSU.TB.02426 CSU.TB.02427 CSU.TB.02428 CSU.TB.02430 CSU.TB.02431
## 1     Incident     Incident     Incident     Incident     Incident     Incident
## 2  1.206122368  1.054281545  1.242026338  1.222782959  1.157624754  1.016192933
## 3  1.192472082  1.155776902  1.140692661  1.144579627  1.097411387  1.034231055
## 4  1.305322575  1.152723272  1.282090163  1.235282459  1.278496113     1.174271
## 5  1.038012976  0.976275507  0.842783709  0.925358784  0.894839488  0.985566735
## 6  1.081272601  1.059776644  1.004844446   1.07542378  0.958268323  0.987615847
##   CSU.TB.02433 CSU.TB.02434 CSU.TB.02435 CSU.TB.02436 CSU.TB.02437 CSU.TB.02442
## 1     Incident     Incident     Incident   Activation     Incident     Incident
## 2   1.02104106   1.03484464  1.271093575  1.219702277  1.085644666   1.12189994
## 3  1.061947108  0.978510199  1.264926698  1.127072933  1.023580199   1.02759182
## 4   1.29247391  1.194677118  1.066420566  1.141932784  1.076097696  1.100008915
## 5  0.911139613  0.938154003  0.979390454  0.937409404  0.935536327  0.701900041
## 6  0.969941401  0.972066008  1.033636026  1.062244238  1.058705014  0.966861954
```

Sinead

```
sinead=read.table('student_dataset_previews/sinead.txt', header=T)

head(sinead)
```

```
##      gt     Area
## 1 ML1>> 2.392419
## 2 ML1>> 1.977150
## 3 ML1>> 2.679428
## 4 ML1>> 2.300645
## 5 ML1>> 2.042630
## 6 ML1>> 2.330245
```

Susan

The dataset I will be using is a subset of data we collect that records the activity of a group of 6 mice every minute for 10-11 days. This data is outputted in an Excel data sheet. To process the data many of the columns are deleted and just tabs with mouse parameters are included.

```
susan_data = read.table("student_dataset_previews/susan.csv", header=T, sep=',', skip=1)
head(susan_data)
```

```
##   StartDate   StartTime    ElapsedTime  Temp Activity Temp.1 Activity.1 Temp.2
## 1    8/6/24 12:00:00 PM 0001:38:52.000 37.35     0.17  38.28       0.40  36.54
## 2    8/6/24 12:01:00 PM 0001:39:52.000 42.25     0.22  37.37       0.17  36.52
## 3    8/6/24 12:02:00 PM 0001:40:52.000 37.47     0.05  43.85       0.20  36.54
## 4    8/6/24 12:03:00 PM 0001:41:52.000 41.15     0.13  37.46       0.10  36.73
## 5    8/6/24 12:04:00 PM 0001:42:52.000 41.06     0.40  37.71       0.17  36.68
```

```
## 6    8/7/24 12:00:00 PM 0025:38:52.000 34.76     0.03  35.43       0.00  36.67
##   Activity.2 Temp.3 Activity.3 Temp.4 Activity.4 Temp.5 Activity.5
## 1       0.33  37.03       0.27  36.79       0.17  37.40       0.27
## 2       0.18  36.50       0.32  42.22       0.67  37.30       0.07
## 3       0.42  37.03       0.45  43.02       0.25  39.16       0.18
## 4       0.40  36.54       0.30  38.49       0.57  37.69       0.28
## 5       0.42  36.34       0.27  60.72       0.15  38.44       0.32
## 6       0.37  34.81       0.00  34.88       0.00  35.93       0.08
```

Xuan

The data examine the different types of mutations observed through sequencing in the TRAF3 genes.

```
library(readxl)
xuan_data = read_xlsx("student_dataset_previews/xuan.txt")
head(xuan_data)
```

```
## # A tibble: 6 x 77
##   Hugo_Symbol CancerGeneType OncoKB_Annotated NCBI_Build Chromosome
##   <chr>       <chr>          <chr>            <chr>           <dbl>
## 1 TRAF3       TSG            Yes              CanFam3.1           8
## 2 TRAF3       TSG            Yes              CanFam3.1           8
## 3 TRAF3       TSG            Yes              CanFam3.1           8
## 4 TRAF3       TSG            Yes              CanFam3.1           8
## 5 TRAF3       TSG            Yes              CanFam3.1           8
## 6 TRAF3       TSG            Yes              CanFam3.1           8
## # i 72 more variables: Start_Position <dbl>, End_Position <dbl>, Strand <chr>,
## #   Variant_Classification <chr>, Variant_Type <chr>, Reference_Allele <chr>,
## #   Tumor_Seq_Allele1 <chr>, Tumor_Seq_Allele2 <chr>, dbSNP_RS <chr>,
## #   Tumor_Sample_Barcode <chr>, Matched_Norm_Sample_Barcode <chr>,
## #   Match_Norm_Seq_Allele1 <chr>, Match_Norm_Seq_Allele2 <chr>, HGVSc <chr>,
## #   HGVSp <chr>, HGVSp_Short <chr>, Transcript_ID <chr>, Exon_Number <dttm>,
## #   t_depth <dbl>, t_ref_count <dbl>, t_alt_count <dbl>, n_depth <dbl>, ...
```