512 Project

Nolan Walker

2/17/2021

Introduction

Data Description

We are looking at the mouse cortex protein expression dataset from Higuera C et. al(2015), where 77 cortex proteins were observed in a mouse model of down syndrome. Some mice were treated with metamine, a drug used to treat alzheimers, to show that learning can be recovered in down syndrome mice. Mice were either assigned to learn, where they were given context and then a shock, or not to learn, where they were only given a shock. This gives a total of 8 classes(2X2X2).There were 15 measurements for each protein, 38 control mice, and 34 down syndrome mice. The goal of this study is to see if there is a difference in protein expression levels across the different classes for the various proteins.

Data exploration

```
dim(data)
```

```
## [1] 1080   82
```

```
str(data[-c(2:78)])
```

```
## 'data.frame':    1080 obs. of  5 variables:
##  $ MouseID  : chr  "309_1" "309_2" "309_3" "309_4" ...
##  $ Genotype : chr  "Control" "Control" "Control" "Control" ...
##  $ Treatment: chr  "Memantine" "Memantine" "Memantine" "Memantine" ...
##  $ Behavior : chr  "C/S" "C/S" "C/S" "C/S" ...
##  $ class    : chr  "c-CS-m" "c-CS-m" "c-CS-m" "c-CS-m" ...
```

```
unique(data$class)
```

```
## [1] "c-CS-m" "c-SC-m" "c-CS-s" "c-SC-s" "t-CS-m" "t-SC-m" "t-CS-s" "t-SC-s"
```

```
names(data[c(2:78)])
```

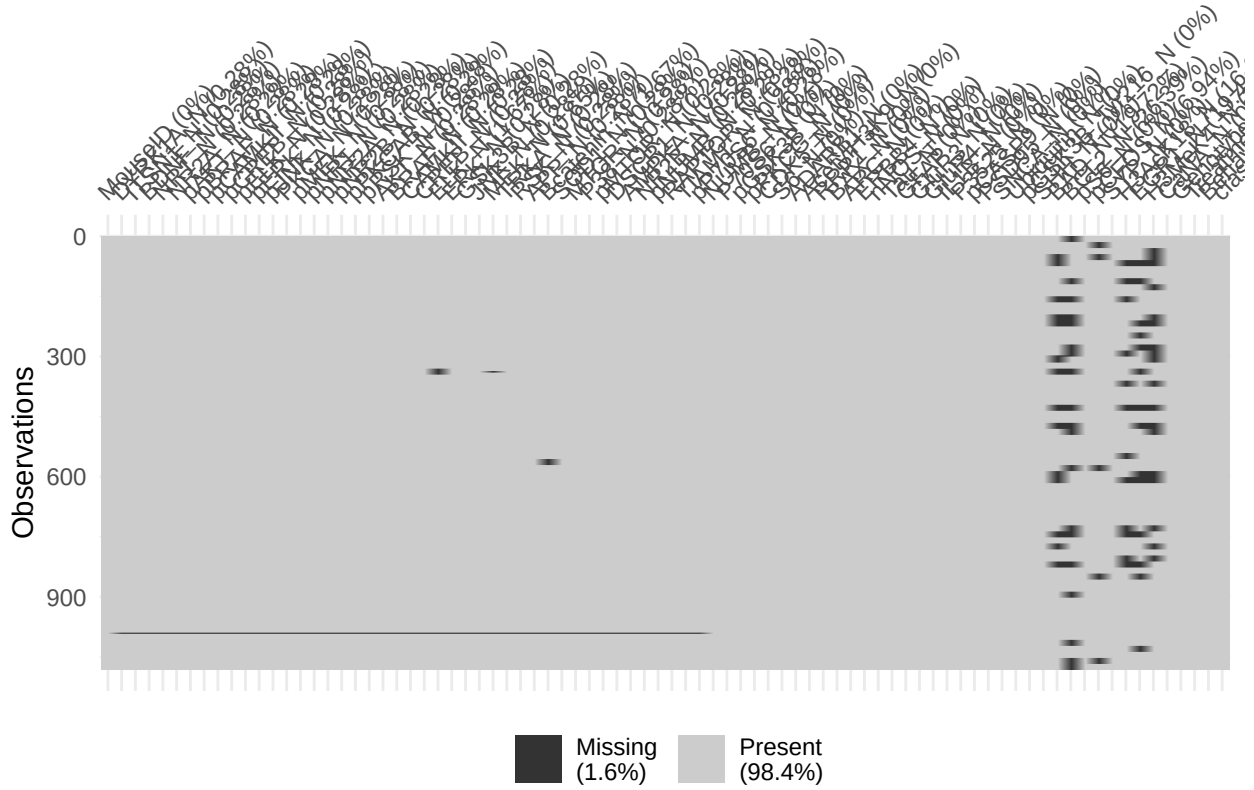
```
##  [1] "DYRK1A_N"        "ITSN1_N"         "BDNF_N"          "NR1_N"
##  [5] "NR2A_N"          "pAKT_N"          "pBRAF_N"         "pCAMKII_N"
##  [9] "pCREB_N"         "pELK_N"          "pERK_N"          "pJNK_N"
## [13] "PKCA_N"          "pMEK_N"          "pNR1_N"          "pNR2A_N"
## [17] "pNR2B_N"         "pPKCAB_N"        "pRSK_N"          "AKT_N"
## [21] "BRAF_N"          "CAMKII_N"        "CREB_N"          "ELK_N"
## [25] "ERK_N"           "GSK3B_N"         "JNK_N"           "MEK_N"
## [29] "TRKA_N"          "RSK_N"           "APP_N"           "Bcatenin_N"
## [33] "SOD1_N"          "MTOR_N"          "P38_N"           "pMTOR_N"
## [37] "DSCR1_N"         "AMPKA_N"         "NR2B_N"          "pNUMB_N"
## [41] "RAPTOR_N"        "TIAM1_N"         "pP70S6_N"        "NUMB_N"
## [45] "P70S6_N"         "pGSK3B_N"        "pPKCG_N"         "CDK5_N"
## [49] "S6_N"            "ADARB1_N"        "AcetylH3K9_N"    "RRP1_N"
## [53] "BAX_N"           "ARC_N"           "ERBB4_N"         "nNOS_N"
```

```
## [57] "Tau_N"           "GFAP_N"          "GluR3_N"         "GluR4_N"
## [61] "IL1B_N"          "P3525_N"         "pCASP9_N"        "PSD95_N"
## [65] "SNCA_N"          "Ubiquitin_N"     "pGSK3B_Tyr216_N" "SHH_N"
## [69] "BAD_N"           "BCL2_N"          "pS6_N"           "pCFOS_N"
## [73] "SYP_N"           "H3AcK18_N"       "EGR1_N"          "H3MeK4_N"
## [77] "CaNA_N"
```

As we can see, there are a lot(1080) of measurements for each of the 77 proteins. There are eight classes; combinations of Genotype, Treatment, and Behavior.

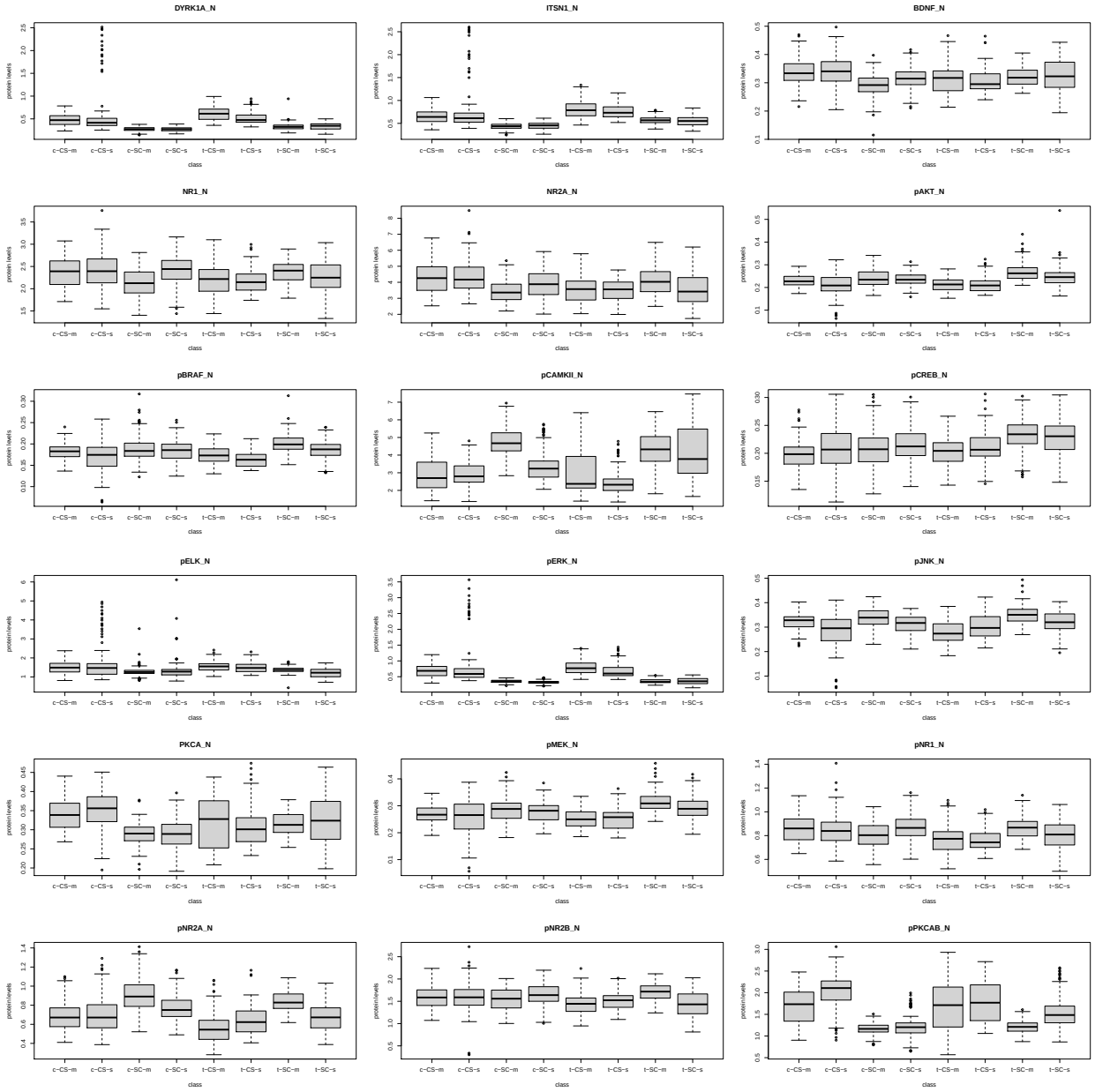
Next to observe missing values.

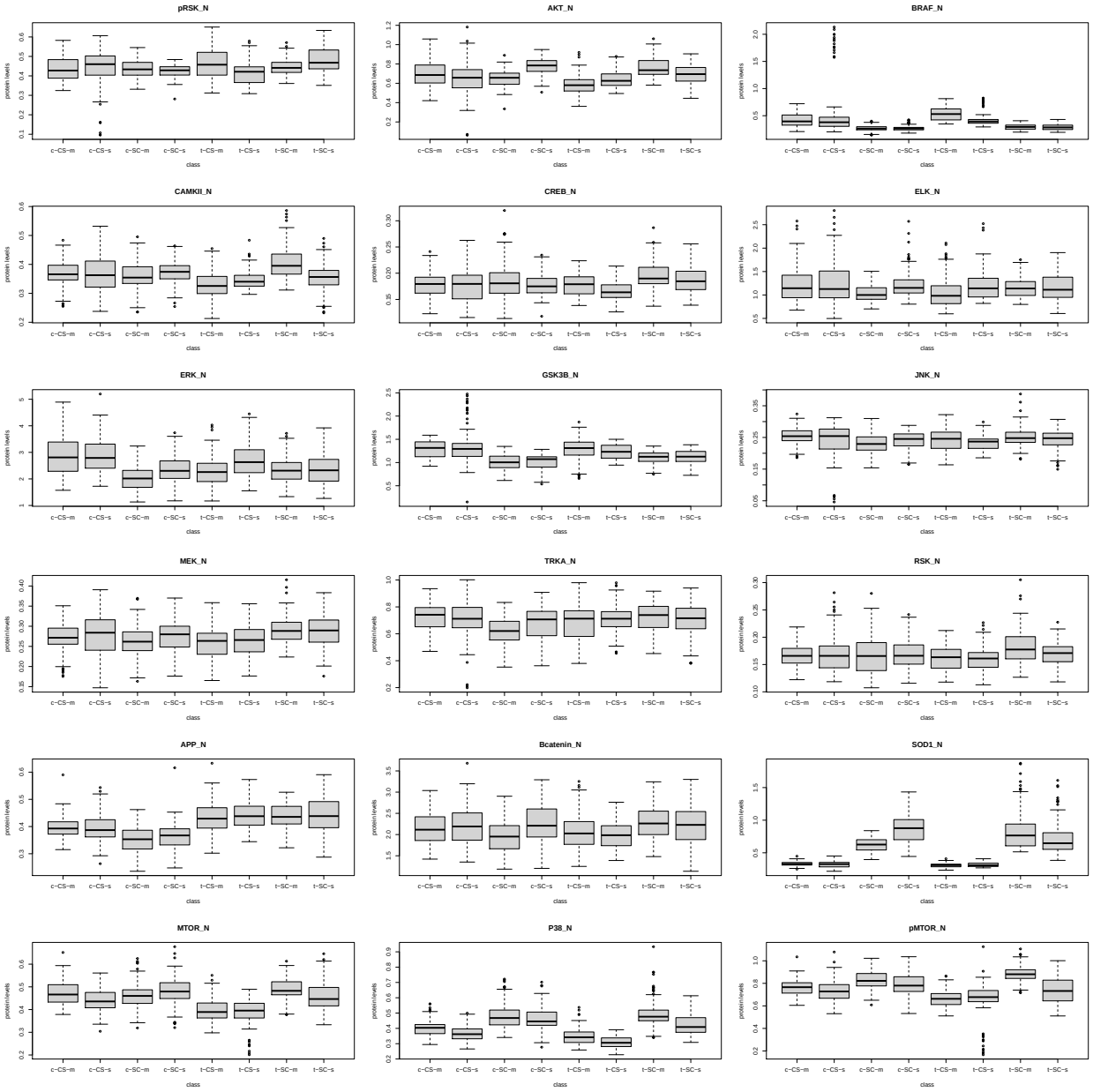
```
vis_miss(data)
```

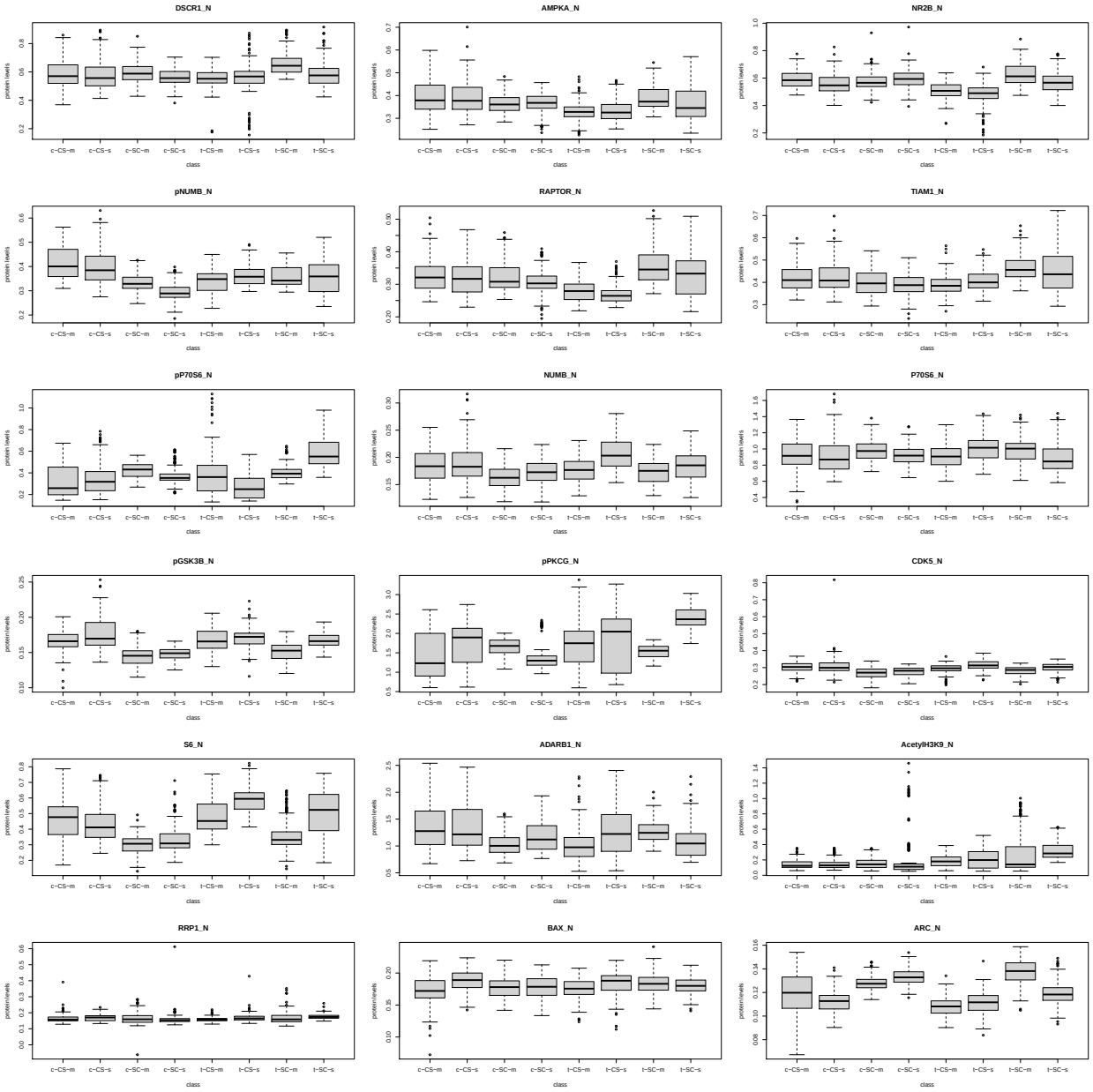


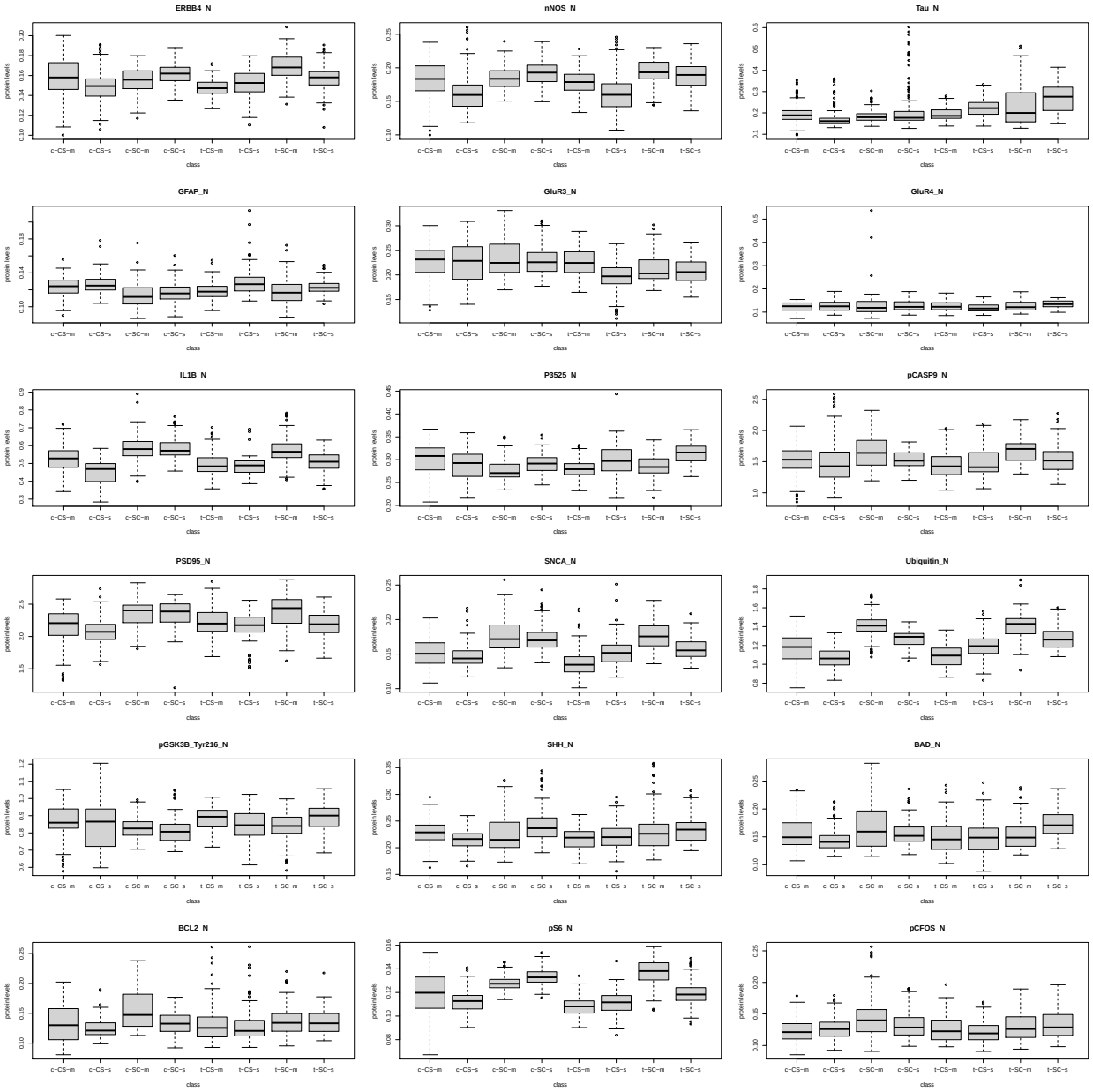
As is observed in the plot of missing values, a few of the proteins towards the right of the plot are missing quite a few observations, so some data imputing must be done.

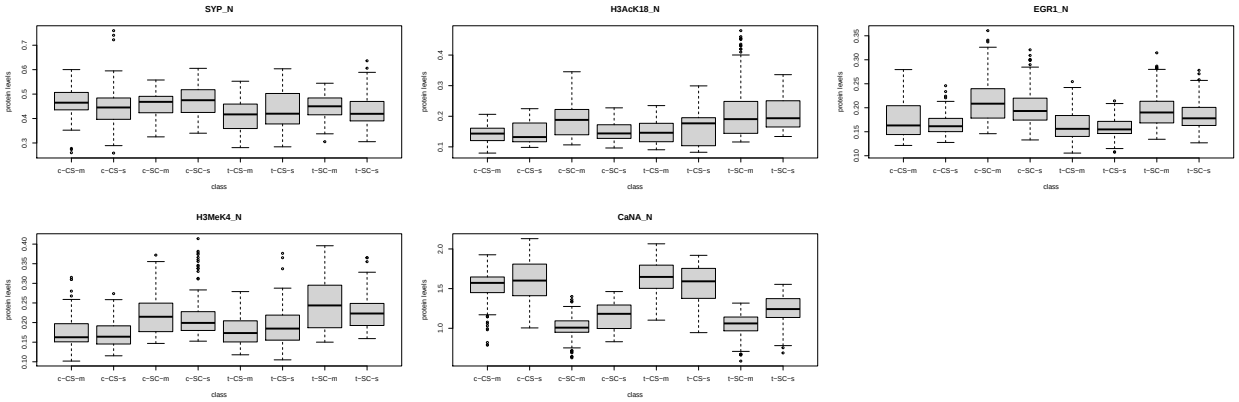
It would be useful to see the distribution of observations for the different proteins.


DYRK1A_N
ITSN1_N
BDNF_N
NR1_N
NR2A_N
pAKT_N
pBRAF_N
pCAMKII_N
pCREB_N
pELK_N
pERK_N
pJNK_N
PKCA_N
pMEK_N
pNR1_N
pNR2A_N
pNR2B_N
pPKCAB_N
protein levels
class
c−CS−m
c−CS−s
c−SC−m
c−SC−s
t−CS−m
t−CS−s
t−SC−m
t−SC−s


pRSK_N
AKT_N
BRAF_N
CAMKII_N
CREB_N
ELK_N
ERK_N
GSK3B_N
JNK_N
MEK_N
TRKA_N
RSK_N
APP_N
Bcatenin_N
SOD1_N
MTOR_N
P38_N
pMTOR_N
protein levels
class
c−CS−m
c−CS−s
c−SC−m
c−SC−s
t−CS−m
t−CS−s
t−SC−m
t−SC−s


DSCR1_N
AMPKA_N
NR2B_N
pNUMB_N
RAPTOR_N
TIAM1_N
pP70S6_N
NUMB_N
P70S6_N
pGSK3B_N
pPKCG_N
CDK5_N
S6_N
ADARB1_N
AcetylH3K9_N
RRP1_N
BAX_N
ARC_N
protein levels
class
c-CS-m
c-CS-s
c-SC-m
c-SC-s
t-CS-m
t-CS-s
t-SC-m
t-SC-s


ERBB4_N
nNOS_N
Tau_N
GFAP_N
GluR3_N
GluR4_N
IL1B_N
P3525_N
pCASP9_N
PSD95_N
SNCA_N
Ubiquitin_N
pGSK3B_Tyr216_N
SHH_N
BAD_N
BCL2_N
pS6_N
pCFOS_N
protein levels
class
c−CS−m
c−CS−s
c−SC−m
c−SC−s
t−CS−m
t−CS−s
t−SC−m
t−SC−s



It looks like some proteins have quite a large difference in measurments across classes. It would make sense to fit multiple mean models to the different proteins to see if there is a difference.

H_null*: there are no diferrences in expression level mean between classes.*

$$H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$$

H_alt*: there is at least one significant difference among the groups.*

$$H_A : \mu_0 \neq \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \neq \mu_7 \neq \mu_8$$

Model fitting

We fit a simple model to each protein :

$\mu\{\text{Protein}|\text{c-CS-m,c-SC-m,c-CS-s,c-SC-s,t-CS-m,t-SC-m,t-CS-s,t-SC-s}\} = \beta_0 + \beta_1 I_{c-SC-m} + \beta_2 I_{c-CS-s} + \beta_3 I_{c-SC-s} + \beta_4 I_{t-CS-m} + \beta_5 I_{t-SC-m} + \beta_6 I_{t-CS-s} + \beta_7 I_{t-SC-s}$

`lm_herb1`:

$$I_{c-SC-m} = \begin{cases} 1 & \text{if class is c-SC-m} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{c-CS-s} = \begin{cases} 1 & \text{if class is c-CS-s} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{c-SC-s} = \begin{cases} 1 & \text{if class is c-SC-s} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{t-CS-m} = \begin{cases} 1 & \text{if class is t-CS-m} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{t-SC-m} = \begin{cases} 1 & \text{if class is t-SC-m} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{t-SC-m} = \begin{cases} 1 & \text{if class is t-SC-m} \\ 0 & \text{otherwise} \end{cases}$$

$$I_{t-SC-s} = \begin{cases} 1 & \text{if class is t-SC-s} \\ 0 & \text{otherwise} \end{cases}$$

```
##            Protein Var_expl       p.value
## 1         DYRK1A_N     0.29  3.980441e-75
## 2          ITSN1_N     0.29  1.230006e-74
## 3           BDNF_N     0.11  8.725596e-24
## 4            NR1_N     0.09  1.357447e-19
## 5           NR2A_N     0.13  1.868668e-28
## 6           pAKT_N     0.19  1.779639e-45
## 7          pBRAF_N     0.16  1.320758e-36
## 8        pCAMKII_N     0.33  2.818994e-88
## 9          pCREB_N     0.11  7.944844e-24
## 10          pELK_N     0.11  9.005484e-23
## 11          pERK_N     0.36  4.569332e-99
## 12          pJNK_N     0.21  6.603188e-51
## 13          PKCA_N     0.16  5.708838e-38
## 14          pMEK_N     0.19  1.157002e-44
## 15          pNR1_N     0.11  5.342219e-25
## 16         pNR2A_N     0.31  5.188787e-82
## 17         pNR2B_N     0.11  1.000318e-22
## 18        pPKCAB_N     0.39 6.452687e-109
## 19          pRSK_N     0.08  3.286824e-16
## 20           AKT_N     0.23  3.370599e-57
## 21          BRAF_N     0.26  4.992471e-65
## 22        CAMKII_N     0.15  1.045206e-33
## 23          CREB_N     0.09  6.970974e-19
## 24           ELK_N     0.06  1.983440e-11
## 25           ERK_N     0.19  3.024075e-44
## 26         GSK3B_N     0.26  2.865826e-65
## 27           JNK_N     0.06  9.683309e-11
## 28           MEK_N     0.07  5.170426e-14
## 29          TRKA_N     0.08  2.068313e-15
## 30           RSK_N     0.05  5.885899e-10
## 31           APP_N     0.33  4.171656e-88
## 32      Bcatenin_N     0.08  1.451842e-15
## 33          SOD1_N     0.66 1.209388e-243
## 34          MTOR_N     0.27  5.161274e-70
## 35           P38_N     0.47 3.545236e-142
## 36         pMTOR_N     0.36  1.039129e-98
## 37         DSCR1_N     0.12  2.887391e-25
## 38         AMPKA_N     0.15  5.249088e-33
## 39          NR2B_N     0.27  2.816364e-68
## 40         pNUMB_N     0.30  4.061030e-80
## 41        RAPTOR_N     0.21  2.472756e-50
## 42         TIAM1_N     0.15  3.951100e-35
## 43        pP70S6_N     0.27  1.437178e-70
## 44          NUMB_N     0.17  6.765151e-40
## 45         P70S6_N     0.06  7.397068e-11
## 46        pGSK3B_N     0.36 7.114063e-100
## 47         pPKCG_N     0.25  1.082395e-62
```

```
## 48          CDK5_N     0.17  8.453641e-39
## 49            S6_N     0.39 1.250355e-110
## 50        ADARB1_N     0.13  2.549053e-30
## 51    AcetylH3K9_N     0.13  7.566911e-29
## 52          RRP1_N     0.03  1.475502e-05
## 53           BAX_N     0.07  6.381704e-14
## 54           ARC_N     0.48 1.728614e-145
## 55         ERBB4_N     0.19  7.883676e-46
## 56          nNOS_N     0.20  8.639413e-47
## 57           Tau_N     0.18  2.985743e-42
## 58          GFAP_N     0.14  3.946997e-32
## 59         GluR3_N     0.11  7.395391e-24
## 60         GluR4_N     0.02  2.040815e-03
## 61          IL1B_N     0.32  2.060474e-84
## 62         P3525_N     0.14  1.033929e-32
## 63        pCASP9_N     0.08  5.008883e-16
## 64         PSD95_N     0.18  1.390435e-41
## 65          SNCA_N     0.33  3.281180e-90
## 66     Ubiquitin_N     0.51 1.628986e-162
## 67 pGSK3B_Tyr216_N     0.07  9.536395e-14
## 68           SHH_N     0.08  1.039664e-17
## 69           BAD_N     0.10  2.102935e-20
## 70          BCL2_N     0.09  1.736128e-19
## 71           pS6_N     0.48 1.728614e-145
## 72         pCFOS_N     0.07  2.001019e-14
## 73           SYP_N     0.08  2.974103e-16
## 74       H3AcK18_N     0.17  1.058385e-38
## 75          EGR1_N     0.21  1.819861e-50
## 76        H3MeK4_N     0.19  7.181599e-44
## 77          CaNA_N     0.61 5.372624e-215
```

```
AOV_df$reject_null <- AOV_df$p.value < alpha
alpha_adj <- alpha / length(unique(Data$class))*(length(unique(Data$class))-1)/2
```