

IvsC_UnivariateAnalysis19Oct2017.R

chanj

Mon Oct 23 13:06:41 2017

```
# Univariate Analyses for the Biochemical Assays of the Ivsc project  
library(ggplot2)  
library(data.table)  
library(scales)  
library(plyr)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':  
##  
##      arrange, count, desc, failwith, id, mutate, rename, summarise,  
##      summarize
```

```
## The following objects are masked from 'package:data.table':  
##  
##      between, first, last
```

```
## The following objects are masked from 'package:stats':  
##  
##      filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

```
library(reshape2)
```

```
##  
## Attaching package: 'reshape2'
```

```
## The following objects are masked from 'package:data.table':  
##  
##      dcast, melt
```

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.4.2
```

```
## Loading required package: Matrix
```

```
library(effects)
```

```
## Loading required package: carData
```

```
## lattice theme set by effectsTheme()  
## See ?effectsTheme for details.
```

```
library(multcomp)
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: survival
```

```
## Loading required package: TH.data
```

```
## Loading required package: MASS
```

```
##  
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      select
```

```
##  
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':  
##  
##      geyser
```

```
library(lmerTest)
```

```
##  
## Attaching package: 'lmerTest'
```

```
## The following object is masked from 'package:lme4':  
##  
##      lmer
```

```
## The following object is masked from 'package:stats':  
##  
##      step
```

```
library(piecewiseSEM)  
library(car)
```

```
##  
## Attaching package: 'car'
```

```
## The following objects are masked from 'package:carData':  
##  
##      Guyer, UN, Vocab
```

```
## The following object is masked from 'package:dplyr':  
##  
##      recode
```

```
library(gridExtra)
```

```
##  
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      combine
```

```
library(cowplot)
```

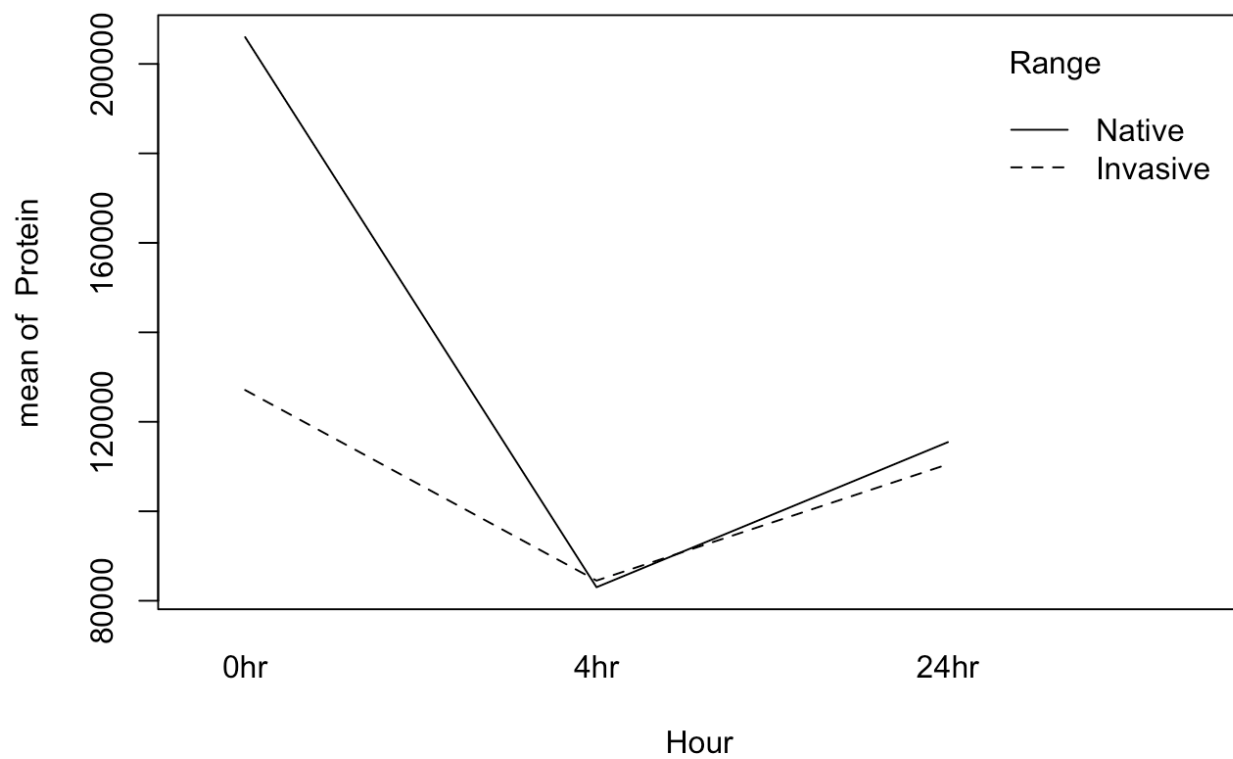
```
##  
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:ggplot2':  
##  
##      ggsave
```

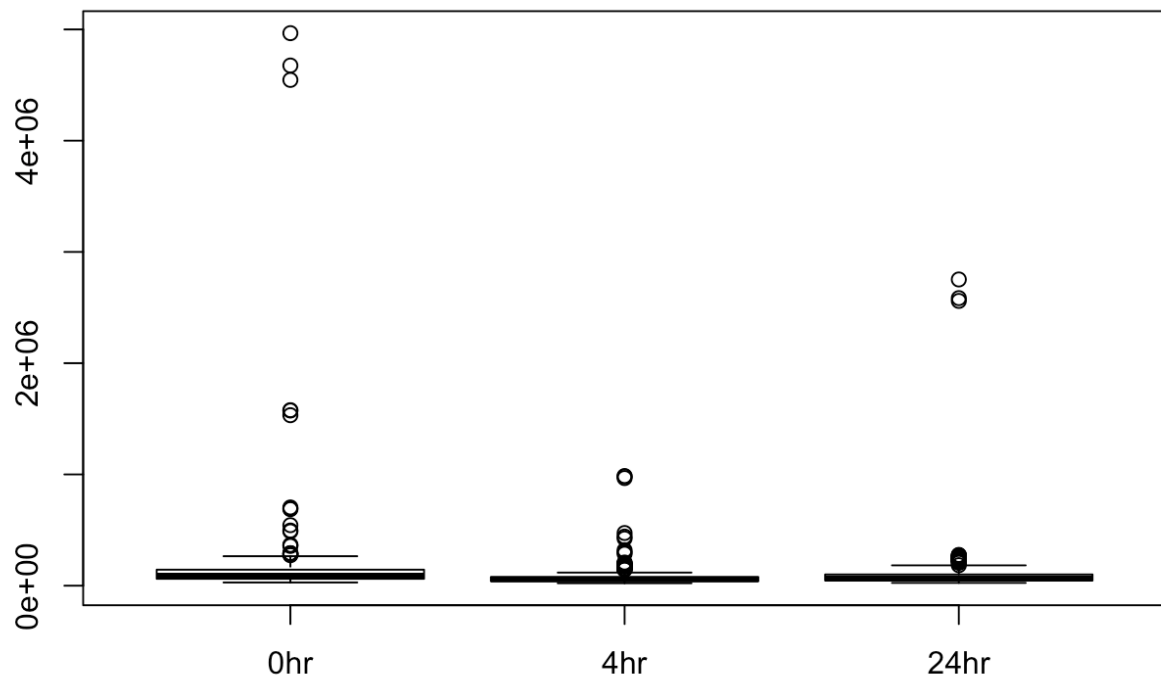
```
library(Rmisc)
```

```
## Loading required package: lattice
```

```
# Data clean up  
# I considered combining data from all three assays and then cleaning but I'm ju  
st going to do it individually. I will combine when I get to multivariate analysi  
s.  
  
# ===== Oct 19, 2017 =====  
setwd("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/Data/Pro  
cessedData/UnivariateAnalyses/")  
  
# The protein, pod, and ppo data files all have similar structure  
  
# =====Protein Quantification =====  
ProteinFile <- read.csv("~/Documents/Friesen lab/MedicagoHerbPopulation/Herbivory  
CollabWSU/Data/ProcessedData/ProteinData22Feb2017.csv")  
  
# Removed unnecessary columns  
ProRemove <- names(ProteinFile) %in% c("PCheck", "Plate", "fileName", "Absorbance  
")  
ProteinFile <- ProteinFile[!ProRemove]  
  
# Removed USDA genotypes  
MPmaster <- read.csv("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCol  
labWSU/Copy of WSUV_MpolMlup_BulkingDesign_3.2016_masterfile -1.csv")  
  
MP_USDA <- MPmaster$Unique.ID[MPmaster$Pod.Produced != "USDA"]  
  
ProteinFile <- ProteinFile[ProteinFile$Sample %in% MP_USDA,]  
ProteinFile$Hour <- factor(ProteinFile$Hour, levels = c("0hr", "4hr", "24hr"))  
  
with(ProteinFile, interaction.plot(Hour, Range, Protein))
```

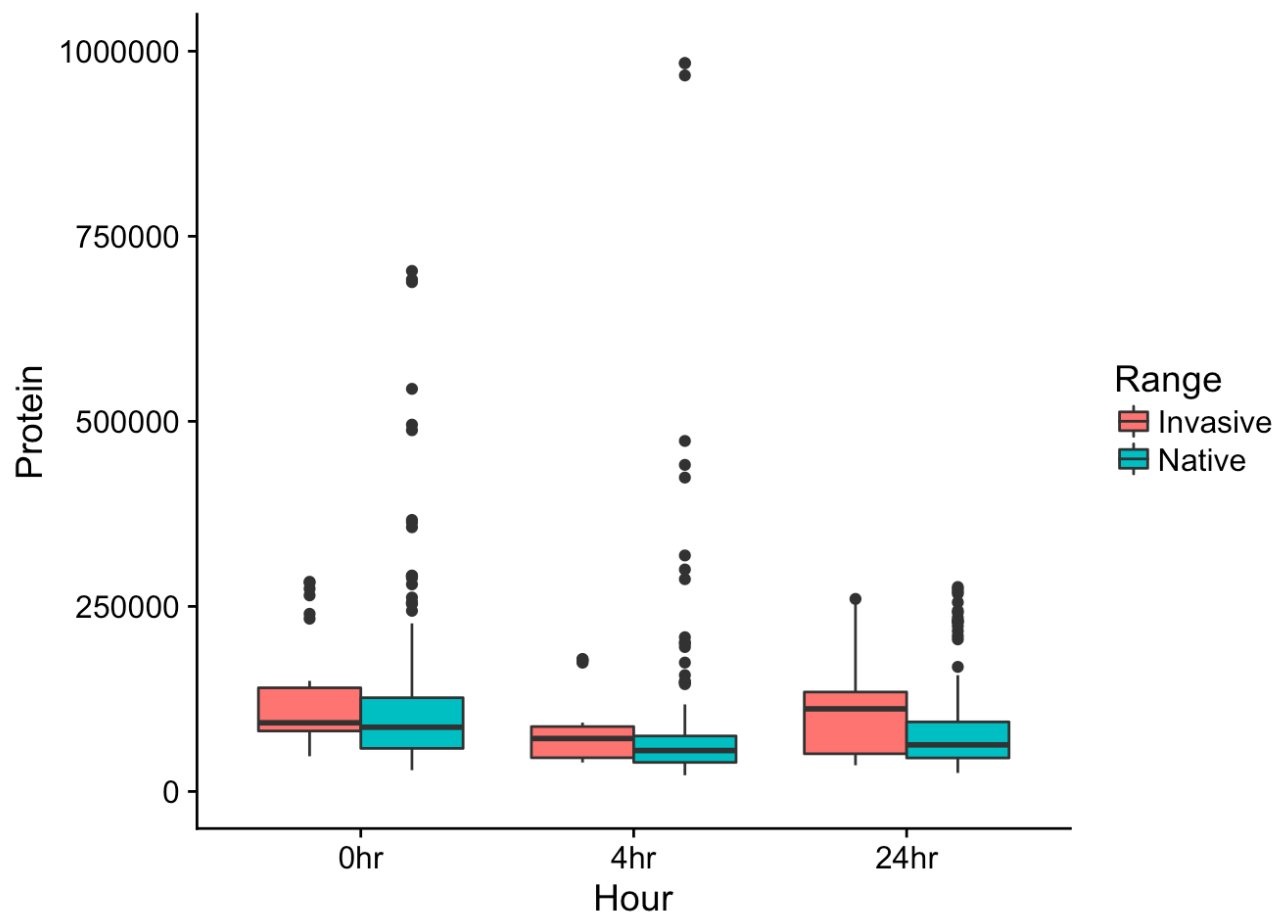


```
boxplot(Protein~Hour, data = ProteinFile)
```



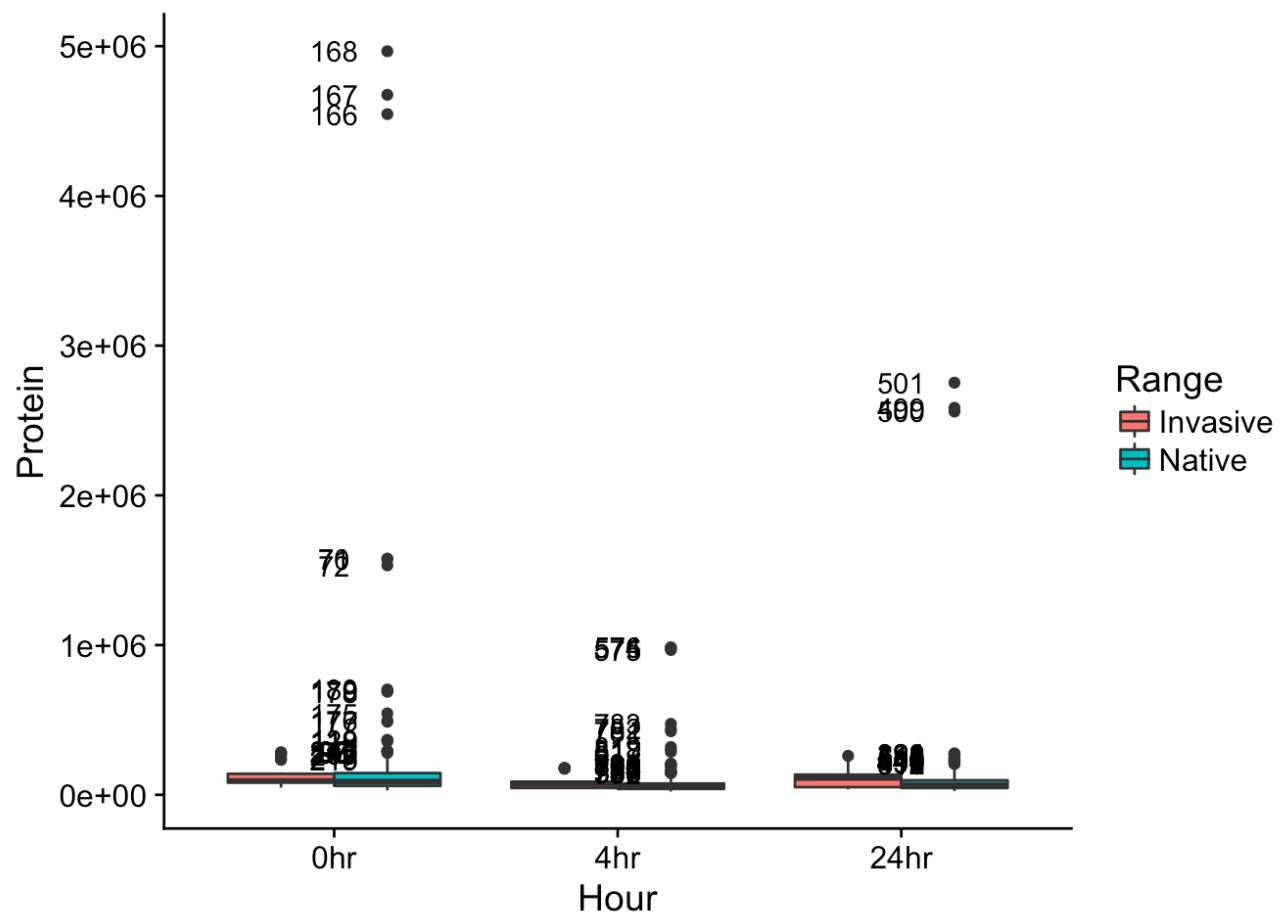
```
ggplot(ProteinFile, aes(Hour, Protein, fill = Range)) + geom_boxplot() +ylim(0, 1  
*10^6)# + facet_wrap(~ Site, scales = "free")
```

```
## Warning: Removed 9 rows containing non-finite values (stat_boxplot).
```



```
# ==== List of outliers ====
is_outlier <- function(x) {
  return(x < quantile(x, 0.25) - 1.5 * IQR(x) | x > quantile(x, 0.75) + 1.5 * IQR
(x))
}

dat <- ProteinFile %>% tibble::rownames_to_column(var="outlier") %>% group_by(Ran
ge, Hour) %>% mutate(is_outlier=ifelse(is_outlier(Protein), Protein, as.numeric(N
A)))
dat$outlier[which(is.na(dat$is_outlier))] <- as.numeric(NA)
ggplot(dat, aes(y=Protein, x=Hour, fill= Range)) + geom_boxplot() + geom_text(aes
(label=outlier), na.rm=TRUE, nudge_y=0.05)
```

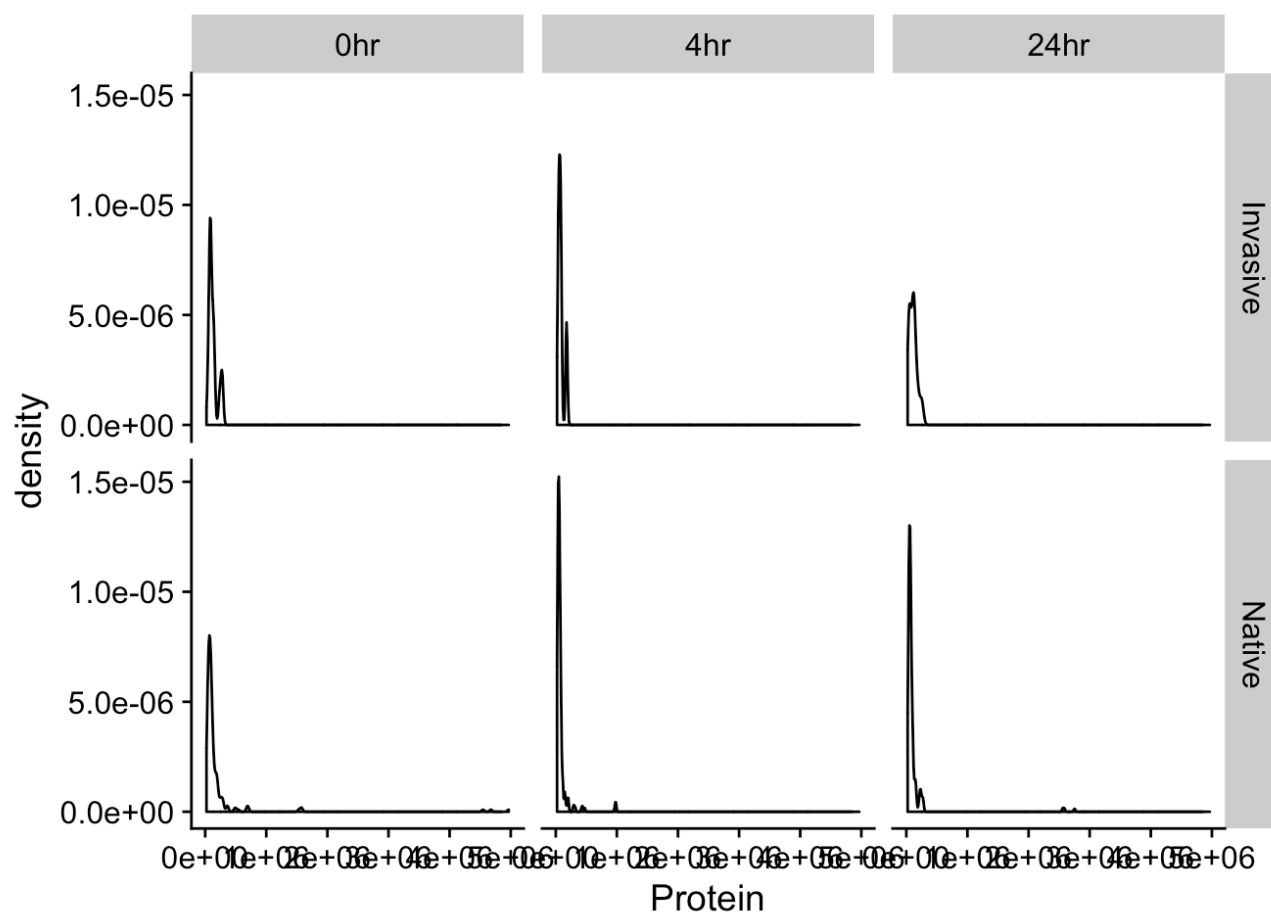


```
ProtOutlier <- subset(dat, !is.na(outlier))
```

```
P1 <- ggplot(ProteinFile, aes(Hour, Protein, fill = Range)) + geom_boxplot()
P2 <- ggplot(ProteinFile, aes(Hour, Protein, fill = Range)) + geom_boxplot() + ylim(0, 1*10^6)
ProBoxplot <- ggdraw() + draw_plot(P1 + theme(legend.position = "none"), 0, 0, 1, 1) + draw_plot(P2, 0.5, 0.6, 0.5, 0.5)
```

```
## Warning: Removed 9 rows containing non-finite values (stat_boxplot).
```

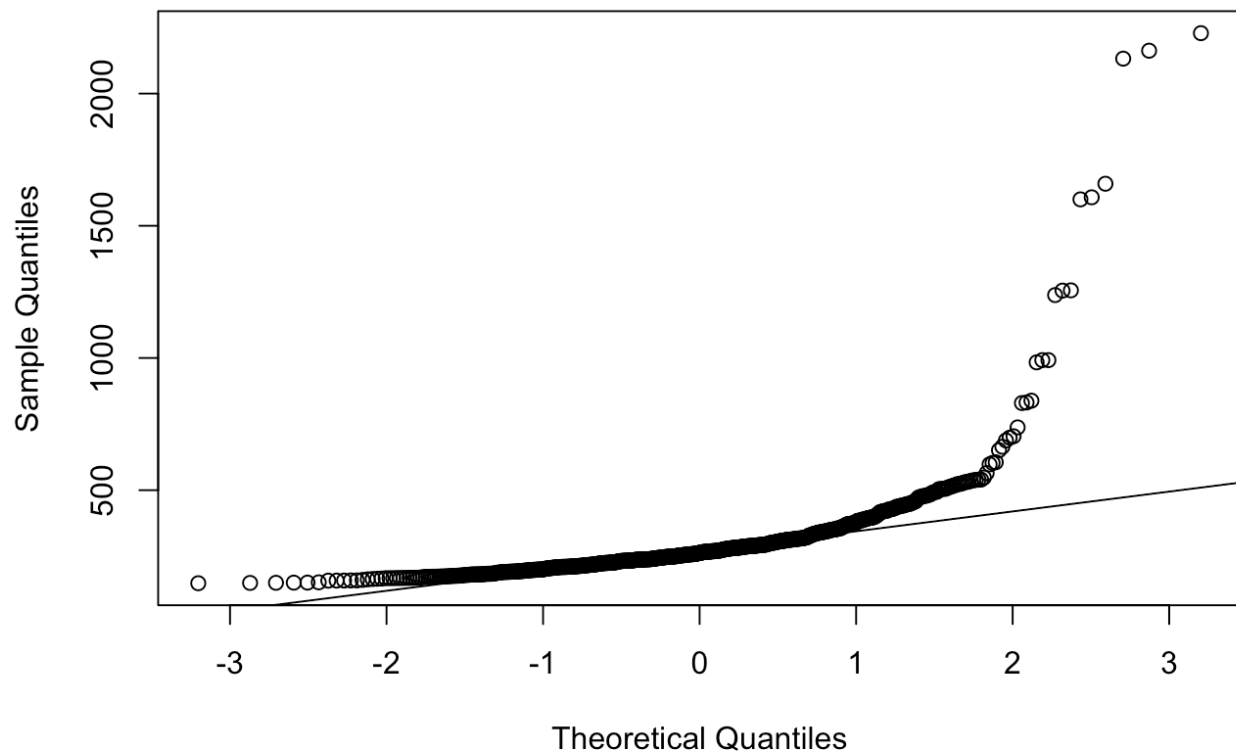
```
save_plot("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/results/exploratory/ProteinBoxplots19Oct2017.pdf", ProBoxplot, base_width = 11, base_height = 8)
ggplot(aes(Protein), data = ProteinFile) + geom_density() + facet_grid(Range ~ Hour)
```

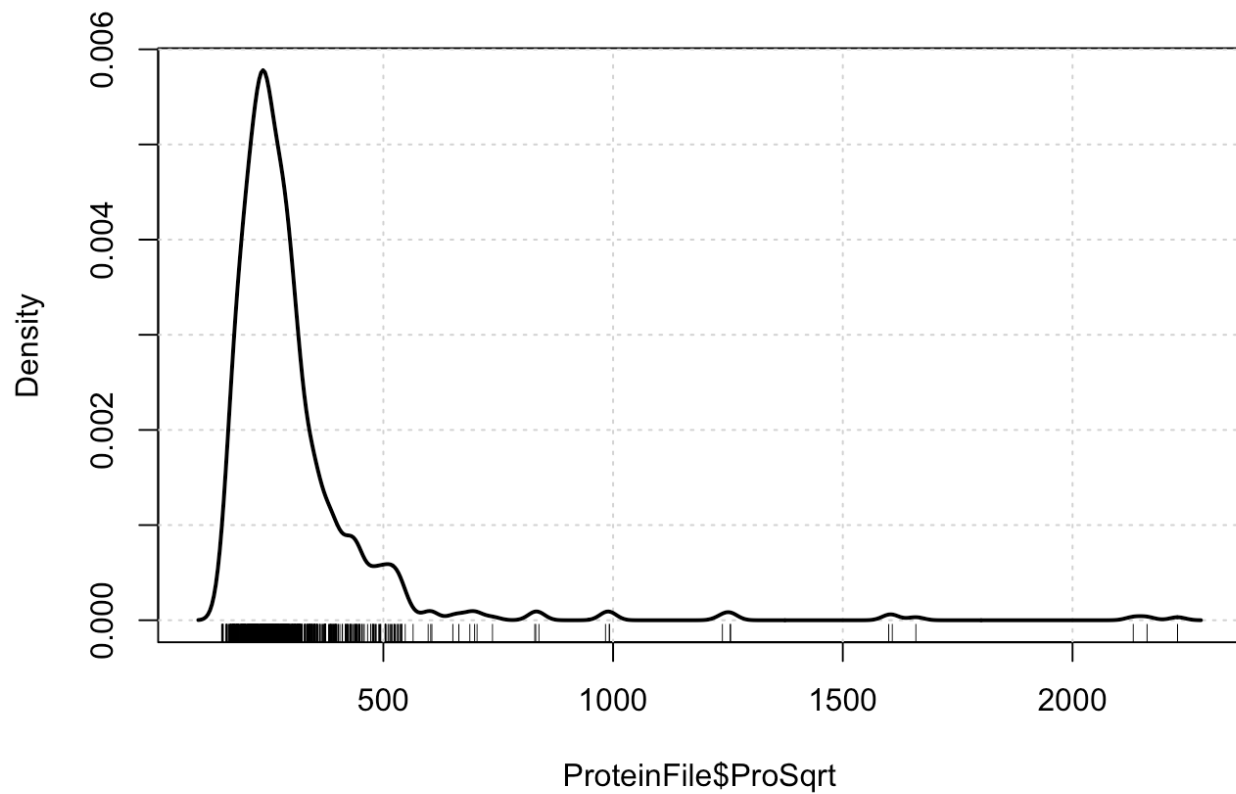
```
ggsave("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/results
/exploratory/ProteinDensityFacetOct2017.pdf", width = 11, height = 8)
```

```
# ==== Transforming to normalize =====
ProteinFile$ProSqrt <- sqrt(ProteinFile$Protein)
qqnorm(ProteinFile$ProSqrt)
qqline(ProteinFile$ProSqrt) # Did not change much
```

Normal Q-Q Plot

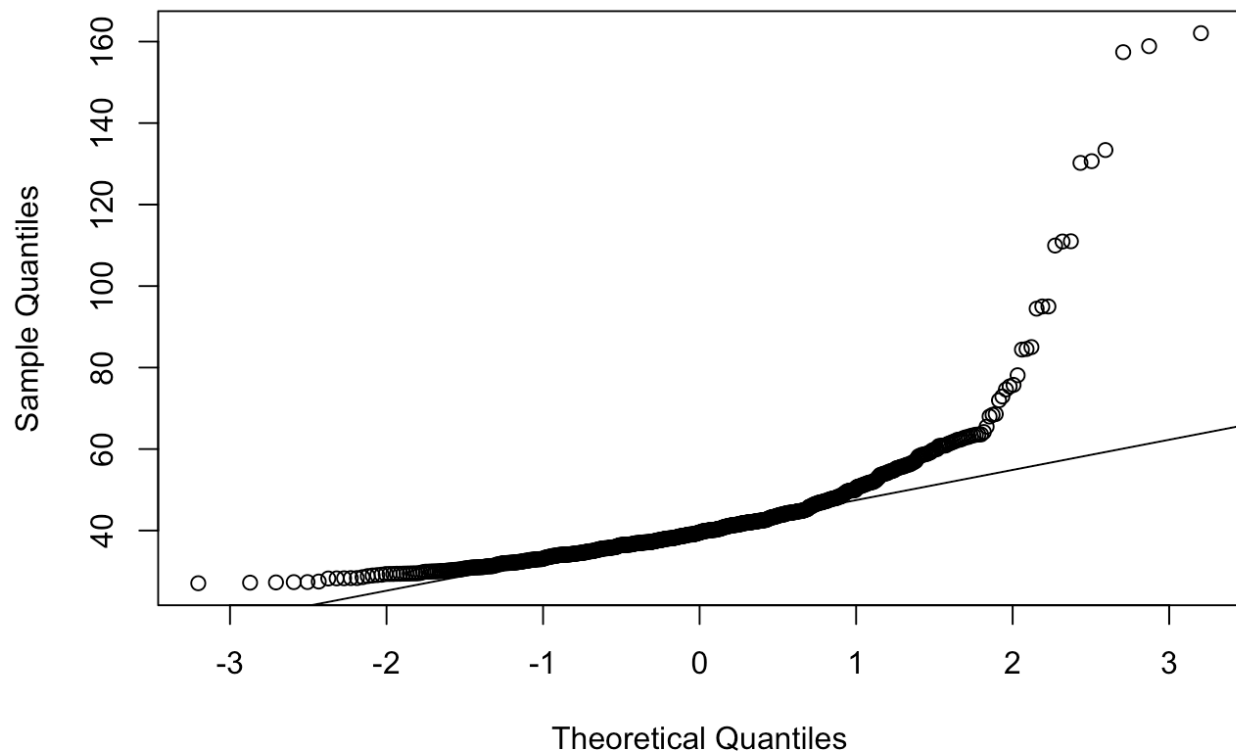


```
densityPlot(ProteinFile$ProSqrt)
```

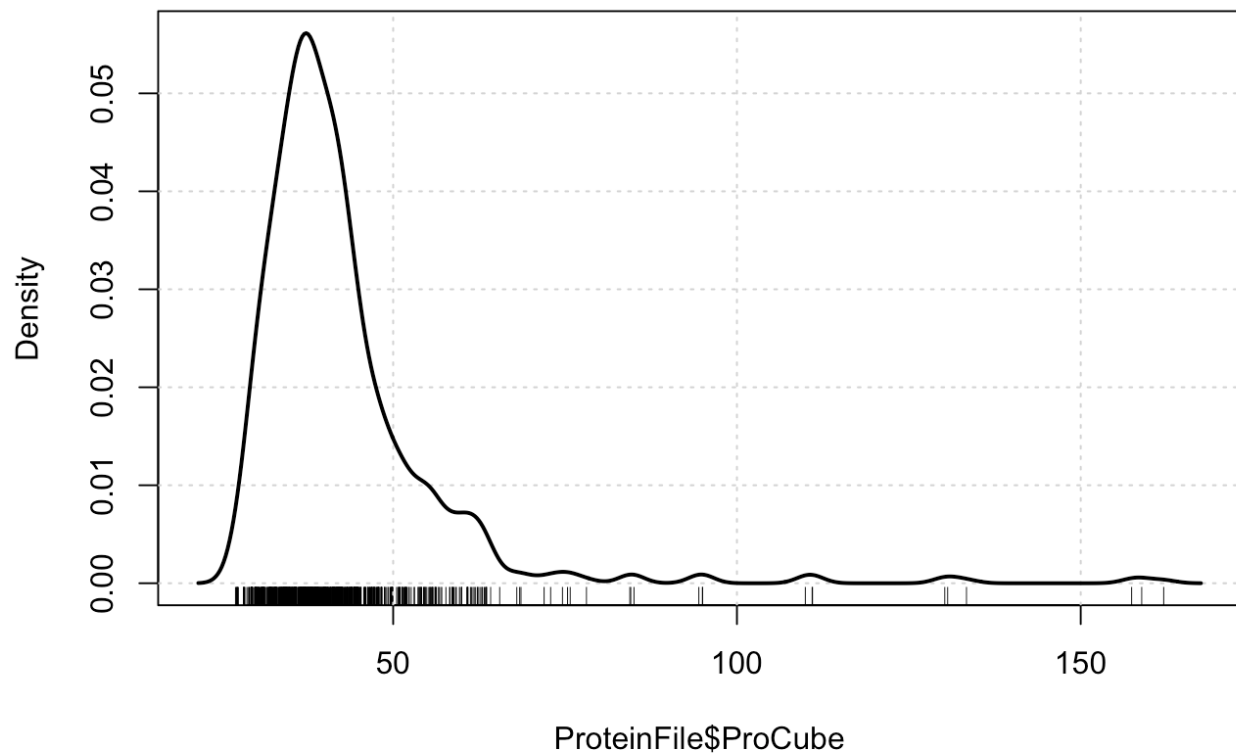


```
ProteinFile$ProCube <- ProteinFile$Protein ^ 0.33  
qqnorm(ProteinFile$ProCube)  
qqline(ProteinFile$ProCube)
```

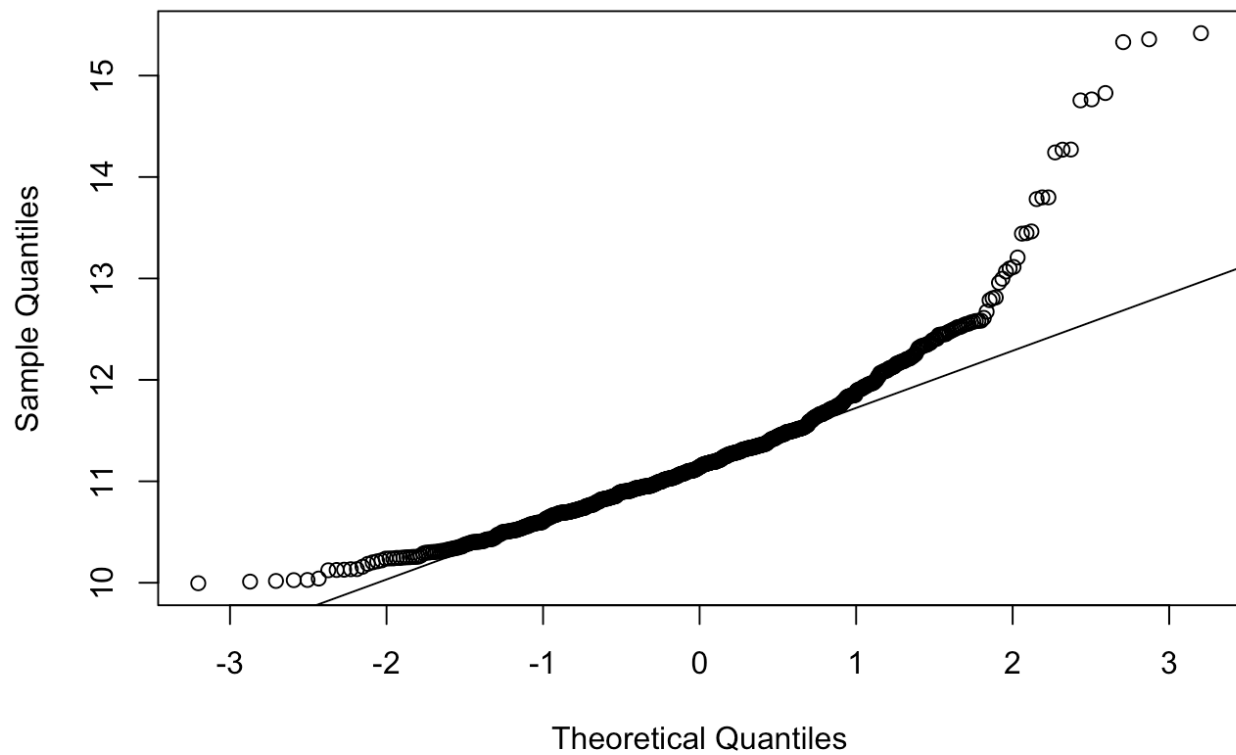
Normal Q-Q Plot



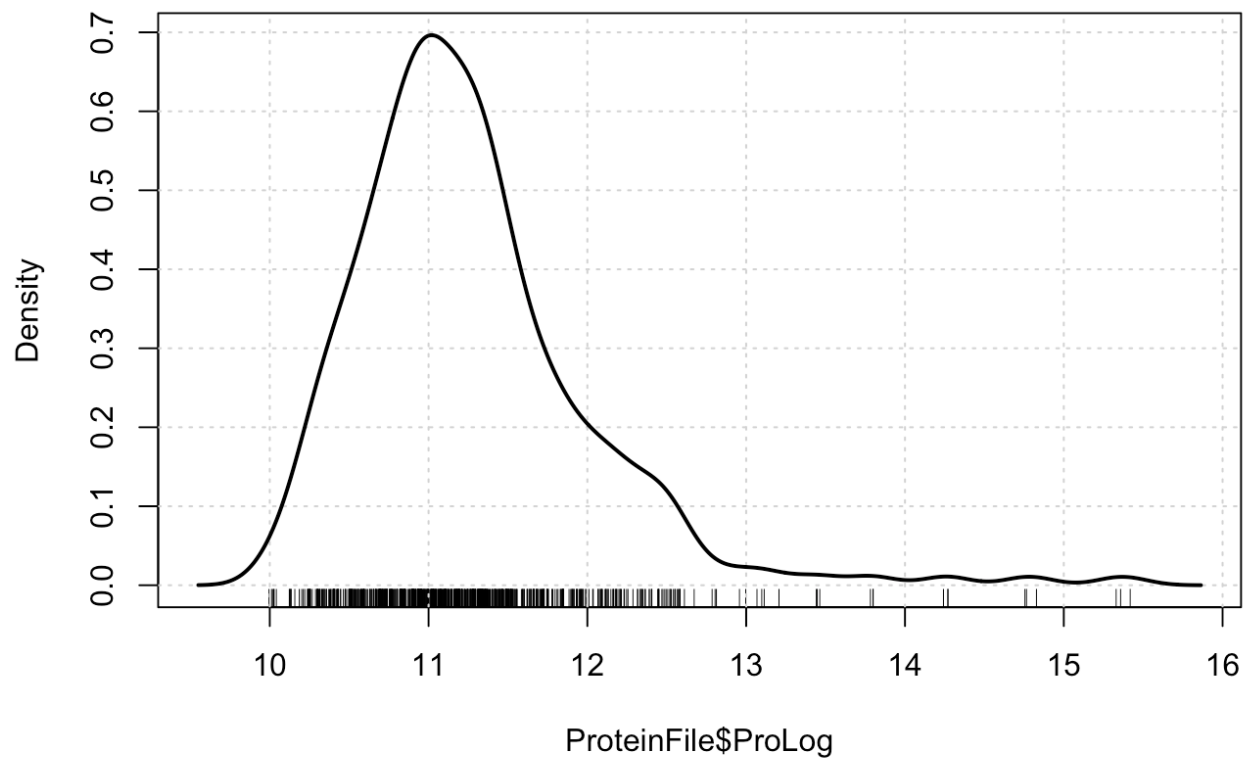
```
densityPlot(ProteinFile$ProCube)
```



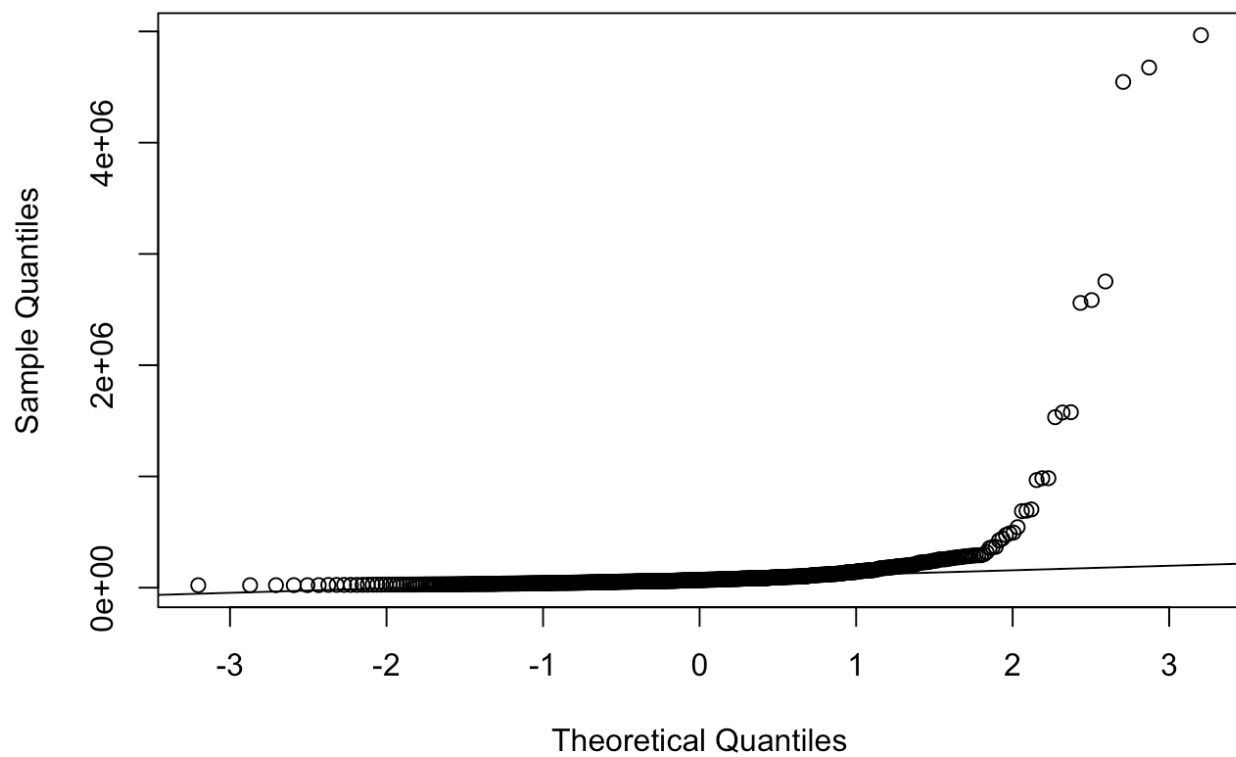
```
ProteinFile$ProLog <- log(ProteinFile$Protein)
qqnorm(ProteinFile$ProLog)
qqline(ProteinFile$ProLog)
```

Normal Q-Q Plot

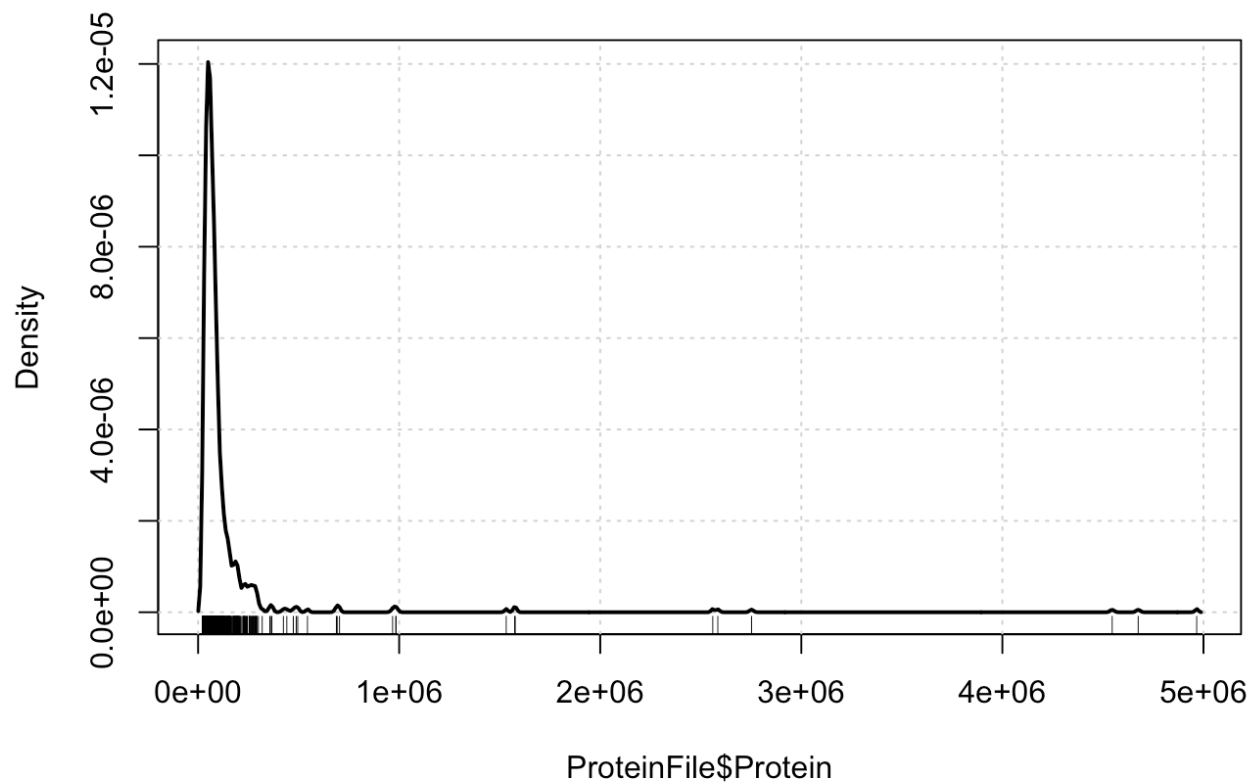
```
densityPlot(ProteinFile$ProLog)
```



```
qqnorm(ProteinFile$Protein)
qqline(ProteinFile$Protein)
```

Normal Q-Q Plot

```
densityPlot(ProteinFile$Protein)
```

```
pdf("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/results/exploratory/ProteinHistTrans19Oct2017.pdf")
par(mfrow=c(4,1))
densityPlot(ProteinFile$Protein)
densityPlot(ProteinFile$ProSqrt)
densityPlot(ProteinFile$ProCube)
densityPlot(ProteinFile$ProLog)
dev.off()
```

```
## quartz_off_screen
##                2
```

```
# will use log transformation
```

```
# Because this is a repeated measures (Time) model, can use a MANOVA for this
```

```
ProteinWide <- dcast(ProteinFile, Sample + Site + Replicate + Genotype + Range ~  
Hour, value.var = "ProLog")
```

```
ProteinWide <- na.exclude(ProteinWide)
```

```
colnames(ProteinWide)[colnames(ProteinWide) == "0hr"] <- "ProteinHr0"
```

```
colnames(ProteinWide)[colnames(ProteinWide) == "4hr"] <- "ProteinHr4"
```

```
colnames(ProteinWide)[colnames(ProteinWide) == "24hr"] <- "ProteinHr24"
```

```
ProMod1 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24) ~ Range, data = ProteinW  
ide)
```

```
blockD <- factor(c("ProteinHr0", "ProteinHr4", "ProteinHr24"), ordered = T)
```

```
blockD <- ordered(blockD, levels = c("ProteinHr0", "ProteinHr4", "ProteinHr24"))
```

```
blockD <- data.frame(blockD)
```

```
Anova(ProMod1, idata = blockD, idesign = ~ blockD)
```

```
##
```

```
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
```

```
##           Df test stat approx F num Df den Df    Pr(>F)  
## (Intercept)  1  0.99824   136641      1    241 < 2.2e-16 ***  
## Range        1  0.01886        5      1    241  0.03235 *  
## blockD       1  0.24001       38      2    240 4.979e-15 ***  
## Range:blockD 1  0.00388        0      2    240  0.62683
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary>Anova(ProMod1, idata = blockD, idesign = ~ blockD))
```

```
##
## Type II Repeated Measures MANOVA Tests:
##
## -----
##
## Term: (Intercept)
##
## Response transformation matrix:
##      (Intercept)
## ProteinHr0      1
## ProteinHr4      1
## ProteinHr24     1
##
## Sum of squares and products for the hypothesis:
##      (Intercept)
## (Intercept)    277524.2
##
## Multivariate Tests: (Intercept)
##      Df test stat approx F num Df den Df      Pr(>F)
## Pillai      1    0.9982 136641.3      1    241 < 2.22e-16 ***
## Wilks       1    0.0018 136641.3      1    241 < 2.22e-16 ***
## Hotelling-Lawley 1 566.9765 136641.3      1    241 < 2.22e-16 ***
## Roy         1 566.9765 136641.3      1    241 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
##
## Term: Range
##
## Response transformation matrix:
##      (Intercept)
## ProteinHr0      1
## ProteinHr4      1
## ProteinHr24     1
##
## Sum of squares and products for the hypothesis:
##      (Intercept)
## (Intercept)      9.410541
##
## Multivariate Tests: Range
##      Df test stat approx F num Df den Df      Pr(>F)
## Pillai      1 0.0188629 4.633357      1    241 0.032349 *
## Wilks       1 0.9811371 4.633357      1    241 0.032349 *
## Hotelling-Lawley 1 0.0192255 4.633357      1    241 0.032349 *
## Roy         1 0.0192255 4.633357      1    241 0.032349 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## -----
##
## Term: blockD
##
## Response transformation matrix:
##           blockD.L  blockD.Q
## ProteinHr0  -7.071068e-01  0.4082483
## ProteinHr4  -7.850462e-17 -0.8164966
## ProteinHr24  7.071068e-01  0.4082483
##
## Sum of squares and products for the hypothesis:
##           blockD.L  blockD.Q
## blockD.L  13.80218 -16.52751
## blockD.Q -16.52751  19.79098
##
## Multivariate Tests: blockD
##           Df test stat approx F num Df den Df    Pr(>F)
## Pillai          1 0.2400072 37.89624      2    240 4.9789e-15 ***
## Wilks           1 0.7599928 37.89624      2    240 4.9789e-15 ***
## Hotelling-Lawley 1 0.3158020 37.89624      2    240 4.9789e-15 ***
## Roy            1 0.3158020 37.89624      2    240 4.9789e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -----
##
## Term: Range:blockD
##
## Response transformation matrix:
##           blockD.L  blockD.Q
## ProteinHr0  -7.071068e-01  0.4082483
## ProteinHr4  -7.850462e-17 -0.8164966
## ProteinHr24  7.071068e-01  0.4082483
##
## Sum of squares and products for the hypothesis:
##           blockD.L  blockD.Q
## blockD.L  0.5071551 -0.10840823
## blockD.Q -0.1084082  0.02317307
##
## Multivariate Tests: Range:blockD
##           Df test stat approx F num Df den Df    Pr(>F)
## Pillai          1 0.0038847 0.4679875      2    240 0.62683
## Wilks           1 0.9961153 0.4679875      2    240 0.62683
## Hotelling-Lawley 1 0.0038999 0.4679875      2    240 0.62683
## Roy            1 0.0038999 0.4679875      2    240 0.62683
##
## Univariate Type II Repeated-Measures ANOVA Assuming Sphericity
##
```

```
##              SS num Df Error SS den Df          F      Pr(>F)
## (Intercept)  92508      1   163.16    241 1.3664e+05 < 2.2e-16 ***
## Range        3         1   163.16    241 4.6334e+00  0.03235 *
## blockD       34         2   211.29    482 3.8317e+01 3.602e-16 ***
## Range:blockD  1         2   211.29    482 6.0490e-01  0.54654
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Mauchly Tests for Sphericity
##
##              Test statistic      p-value
## blockD              0.9167 2.9308e-05
## Range:blockD        0.9167 2.9308e-05
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
## for Departure from Sphericity
##
##              GG eps Pr(>F[GG])
## blockD        0.9231  4.05e-15 ***
## Range:blockD  0.9231    0.5338
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              HF eps  Pr(>F[HF])
## blockD        0.9299056 3.269386e-15
## Range:blockD  0.9299056 5.349466e-01
```

```
# ====POD assay ==== Oct 21, 2017
PODfile <- read.csv("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryColl
abWSU/Data/ProcessedData/WSU_PODfiles11Sept2017.csv")

# Removed unnecessary columns
PODRemove <- names(PODfile) %in% c("PCheck", "Plate", "fileName", "values", "Matc
hValue", "CtrlVal", "ABStrt_ABSctr")
PODfile <- PODfile[!PODRemove]

PODfile <- PODfile[PODfile$Sample %in% MP_USDA,]
PODfile$Hour <- factor(PODfile$Hour, levels = c("0hr", "4hr", "24hr"))
ggplot(PODfile, aes(Hour, AbsFreshWeight, fill = Range)) + geom_boxplot()# +ylim(
0, 1*10^6) + facet_wrap(~ Site, scales = "free")
```

