

multivariate_analysis.R

chanj

Fri Nov 17 09:08:33 2017

```
# Multivariate analysis of biochemical assayss
```

```
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
```

```
library(heplots)  
library(Hmisc)
```

```
## Loading required package: Formula
```

```
##  
## Attaching package: 'Hmisc'
```

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

```
## The following objects are masked from 'package:dplyr':  
##  
##      combine, src, summarize
```

```
## The following objects are masked from 'package:plyr':  
##  
##      is.discrete, summarize
```

```
## The following objects are masked from 'package:base':  
##  
##      format.pval, round.POSIXt, trunc.POSIXt, units
```

```
library(RcmdrMisc)
```

```
## Loading required package: sandwich
```

```
##  
## Attaching package: 'RcmdrMisc'
```

```
## The following object is masked from 'package:Hmisc':  
##  
##      Dotplot
```

```
library(corrplot)
```

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

```
## corrplot 0.84 loaded
```

```
setwd("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/Data/ProcessedData/")

# Here I would read in the three univariate files, but I already have them loaded
. Remember for future runs.

PODWide <- read.csv("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/Data/ProcessedData/PeroxidaselDataWideFormat25Oct2017.csv")

ProteinWide <- read.csv("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/Data/ProcessedData/ProteinWideFormat25Oct2017.csv")

PPOWide <- read.csv("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/Data/ProcessedData/PolyphenolDataWideFormat25Oct2017.csv")

BiocAssaysWide <- merge(ProteinWide, PODWide)
BiocAssaysWide <- merge(BiocAssaysWide, PPOWide)

write.table(BiocAssaysWide, file = "AllBiocAssaysMANOVADataStyle14Nov2017.csv", sep = ",", row = F)

# Time for the big jump!
BCmod1 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24, PODHr0, PODHr4, PODHr24, PPOHr0, PPOHr4, PPOHr24) ~ Range, data = BiocAssaysWide)

Assay <- factor(rep(c("Protein", "POD", "PPO"), each = 3))
MeasTime <- factor(rep(c("Hr0", "Hr4", "Hr24"), 3), levels = c("Hr0", "Hr4", "Hr24"))

BCmod2 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24, PODHr0, PODHr4, PODHr24, PPOHr0, PPOHr4, PPOHr24) ~ Site, data = BiocAssaysWide)

idata <- data.frame(Assay, MeasTime)

anova.BCmod1 <- Anova(BCmod1, idata= idata, idesign = ~ Assay * MeasTime, item = "Assay:MeasTime", type = 3)

anova.BCmod2 <- Anova(BCmod2, idata= idata, idesign = ~ Assay * MeasTime, type = 3)

pdf("~/Documents/Friesen lab/MedicagoHerbPopulation/HerbivoryCollabWSU/results/exploratory/TempHE.pdf")
par(mfrow=c(3,3))
for(i in 1:9){
```

```

for(j in 1:9){
  heplot(BCmod2, variables = c(i,j), cex = 0.5)
}
}
dev.off()

```

```

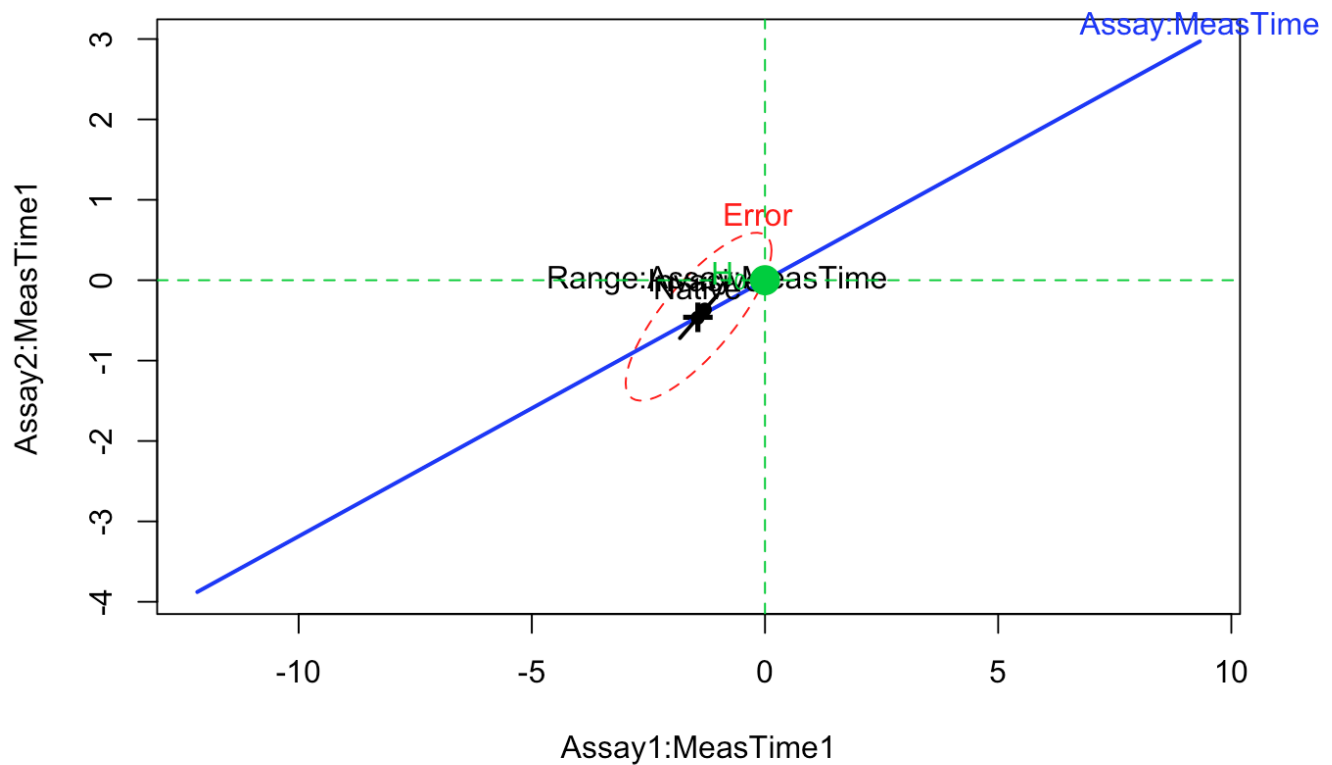
## quartz_off_screen
##                2

```

```

heplot(BCmod1, idata= idata, idesign = ~ Assay * MeasTime, item = "Assay:MeasTime")

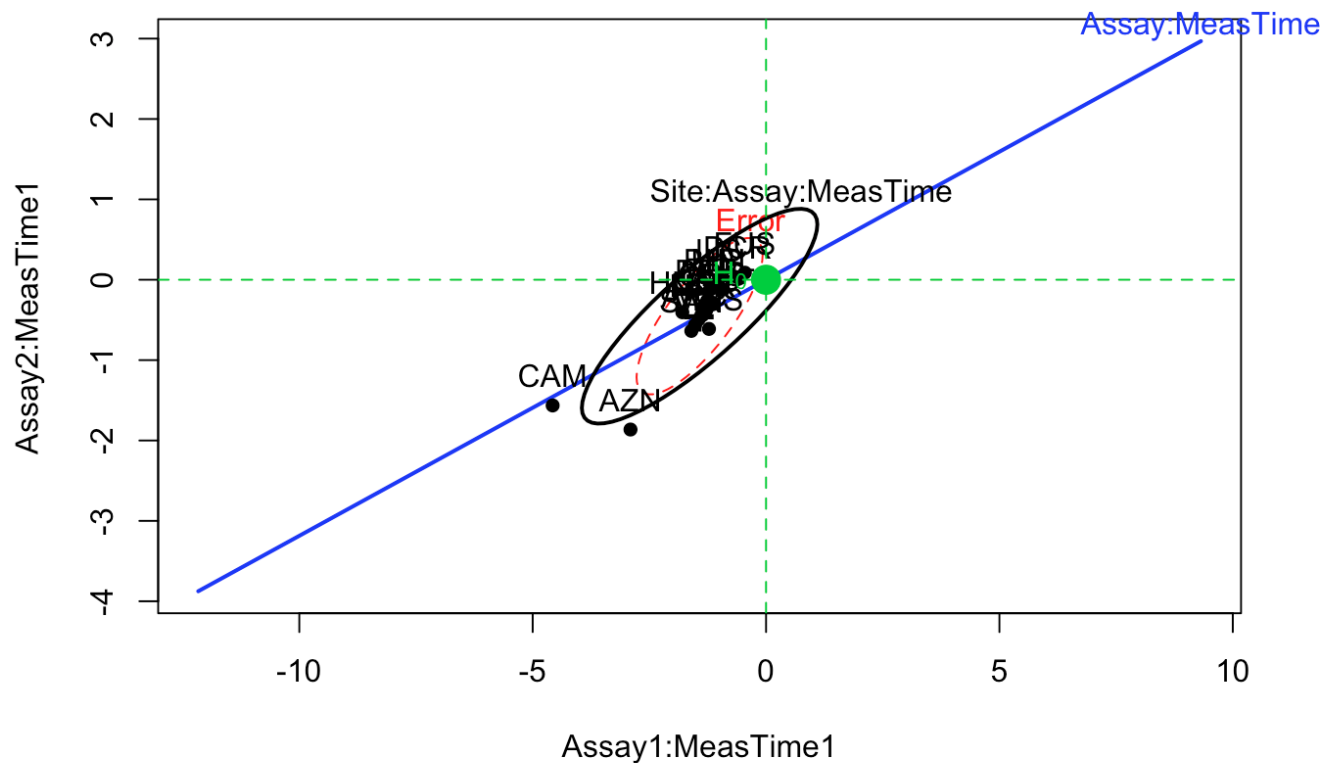
```



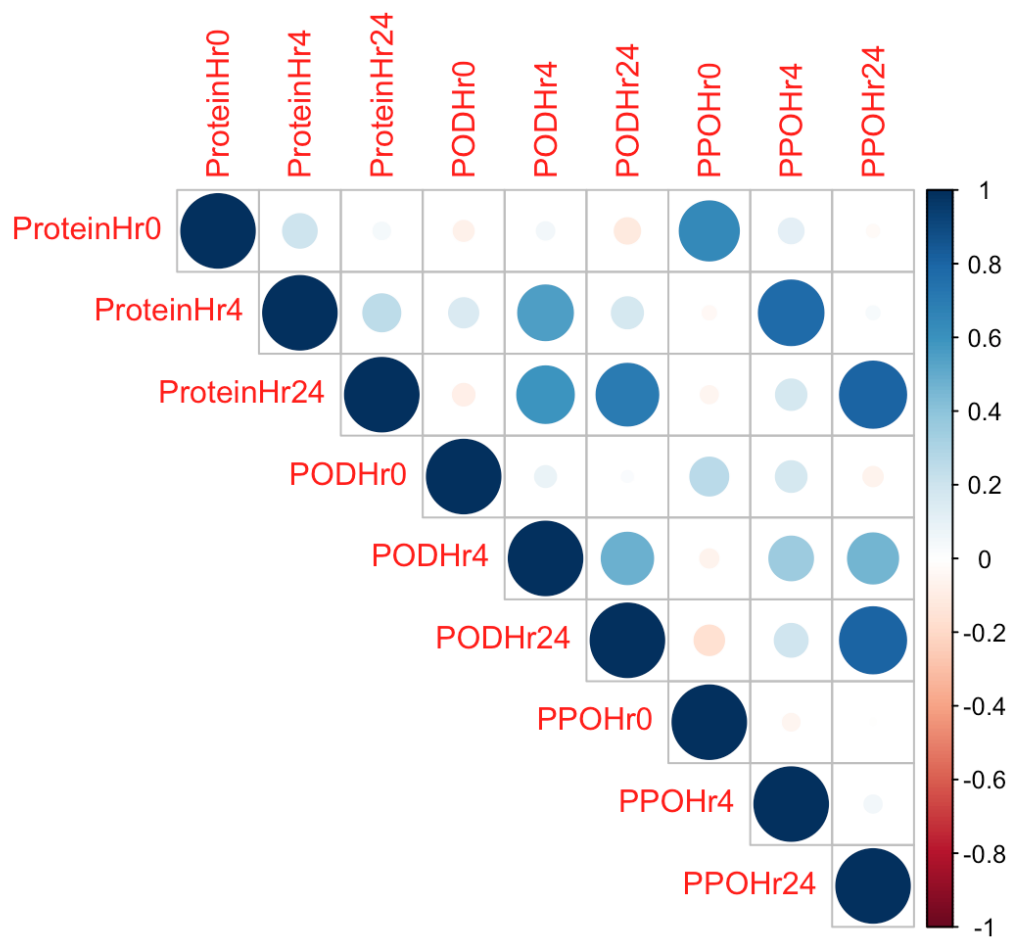
```

heplot(BCmod2, idata= idata, idesign = ~ Assay * MeasTime, item = "Assay:MeasTime")

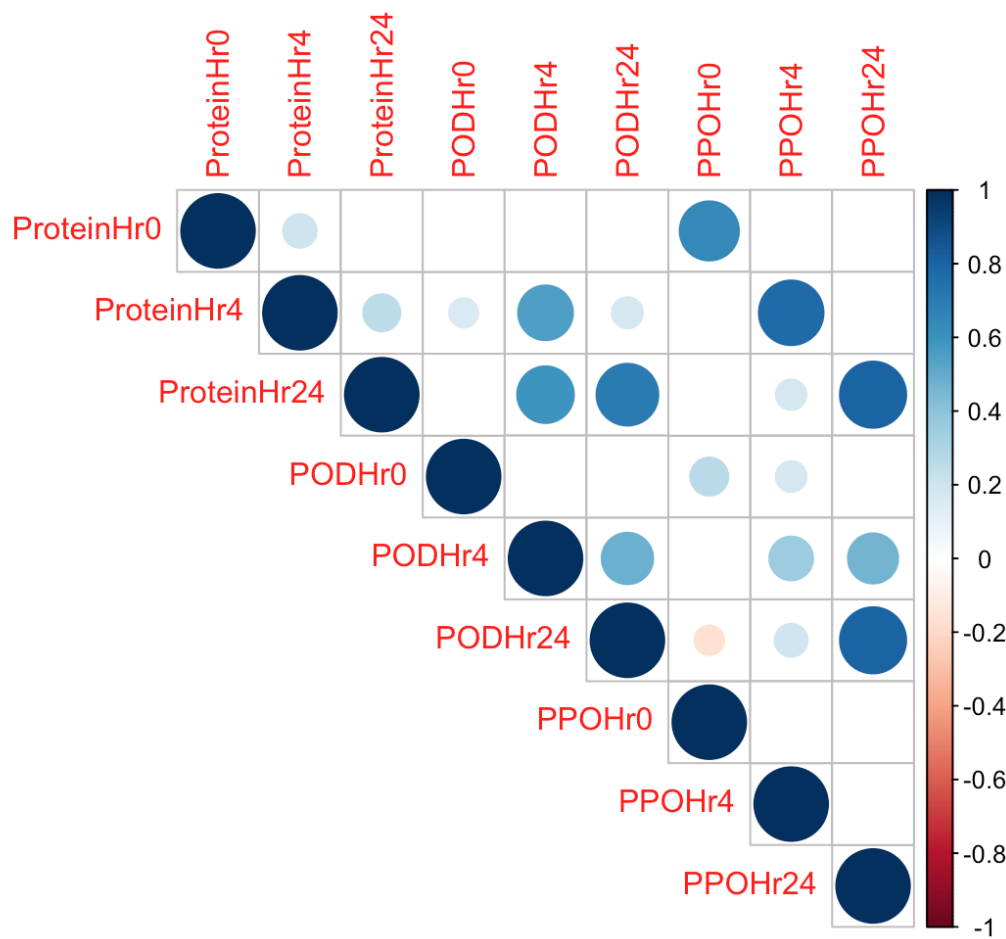
```



```
CorData <- BiocAssaysWide[, c(6:14)]  
  
CDmatrix1 <- cor(CorData)  
CDmatrix2 <- rcorr(as.matrix(CorData))  
CDmatrix3 <- rcorr.adjust(as.matrix(CorData))  
  
corrplot(CDmatrix1, type = "upper")
```

```
corrplot(CDmatrix2$r, type = "upper", p.mat = CDmatrix2$P, sig.level = 0.05, insi
g = "blank")
```



```
corrplot(CDmatrix3$R$r, type = "upper", p.mat = apply(CDmatrix3$P, 2, as.numeric)
, sig.level = 0.05, insig = "blank")
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

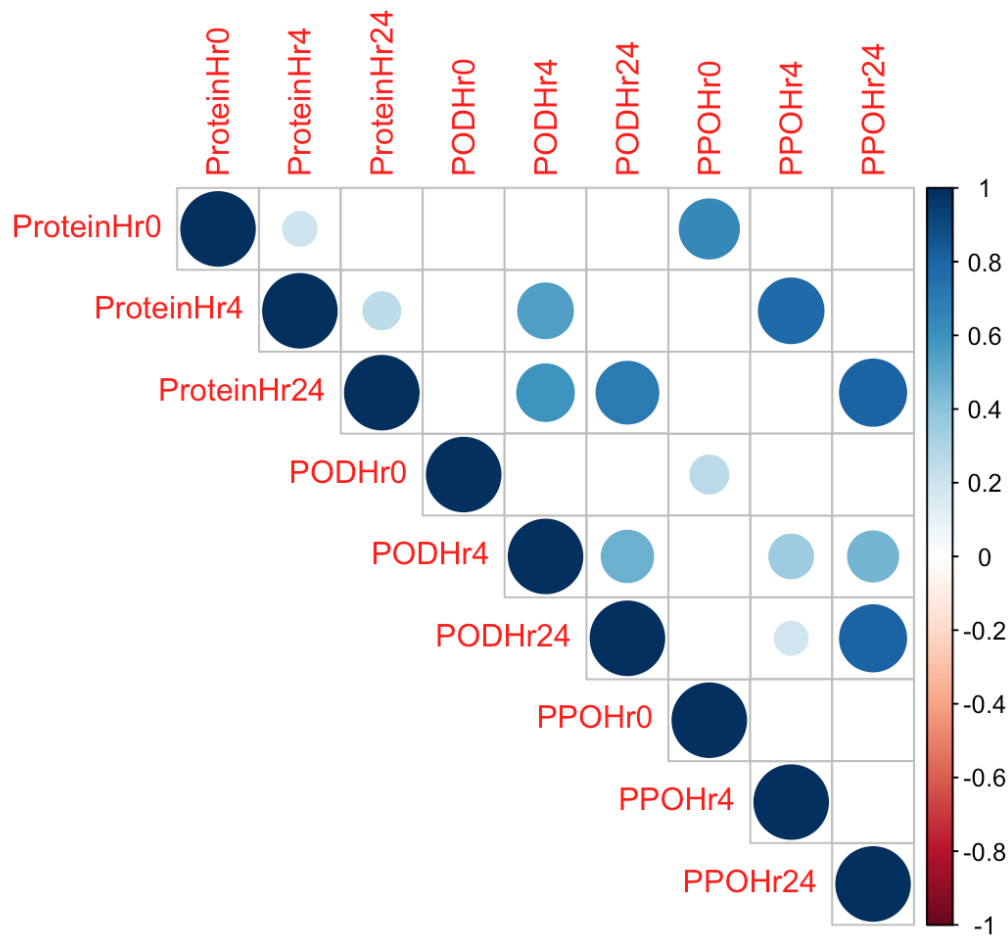
```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(CDmatrix3$P, 2, as.numeric): NAs introduced by coercion
```



```
PvalueAdj <- as.matrix(CDmatrix3$P)
PmatNum <- apply(PvalueAdj, 2, as.numeric)
```

```
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
```

```
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
## Warning in apply(PvalueAdj, 2, as.numeric): NAs introduced by coercion
```

```
#corrplot(CDmatrix$R$r, type = "upper", p.mat = CDmatrix$P.unadjust, sig.level =
0.05, insig = "blank", method = "number")

# ===== Making a figure of the model Nov 14, 2017 =====
# Need to use original data that has not been log transformed
ProteinUntra <- read.csv("/Users/chanj/Documents/Friesen lab/MedicagoHerbPopulati
on/HerbivoryCollabWSU/Data/ProcessedData/ProteinData22Feb2017.csv")

PODUntra <- read.csv("/Users/chanj/Documents/Friesen lab/MedicagoHerbPopulation/H
erbivoryCollabWSU/Data/ProcessedData/WSU_PODfiles11Sept2017.csv")

PPOUntra <- read.csv("/Users/chanj/Documents/Friesen lab/MedicagoHerbPopulation/
HerbivoryCollabWSU/Data/ProcessedData/WSU_PPOfiles11Sept2017.csv")

# Remove all extraneous columns
ProteinUnneeded <- names(ProteinUntra) %in% c("Well", "Weight", "Hour", "PCheck",
"Plate", "fileName", "Absorbance")
ProteinUntra <- ProteinUntra[!ProteinUnneeded]

PODKeep <- names(PODUntra) %in% c("Sample", "Replicate", "Time", "AbsFreshWeight"
, "Range", "Genotype", "Site")
PODUntra <- PODUntra[PODKeep]

PPOKeep <- names(PPOUntra) %in% c("Sample", "Replicate", "Time", "AbsFreshWeight"
, "Range", "Genotype", "Site")
PPOUntra <- PPOUntra[PPOKeep]

# Rename response variable so that can bind the rows
names(ProteinUntra)[names(ProteinUntra) == "Protein"] <- "Response"
names(PODUntra)[names(PODUntra) == "AbsFreshWeight"] <- "Response"
names(PPOUntra)[names(PPOUntra) == "AbsFreshWeight"] <- "Response"

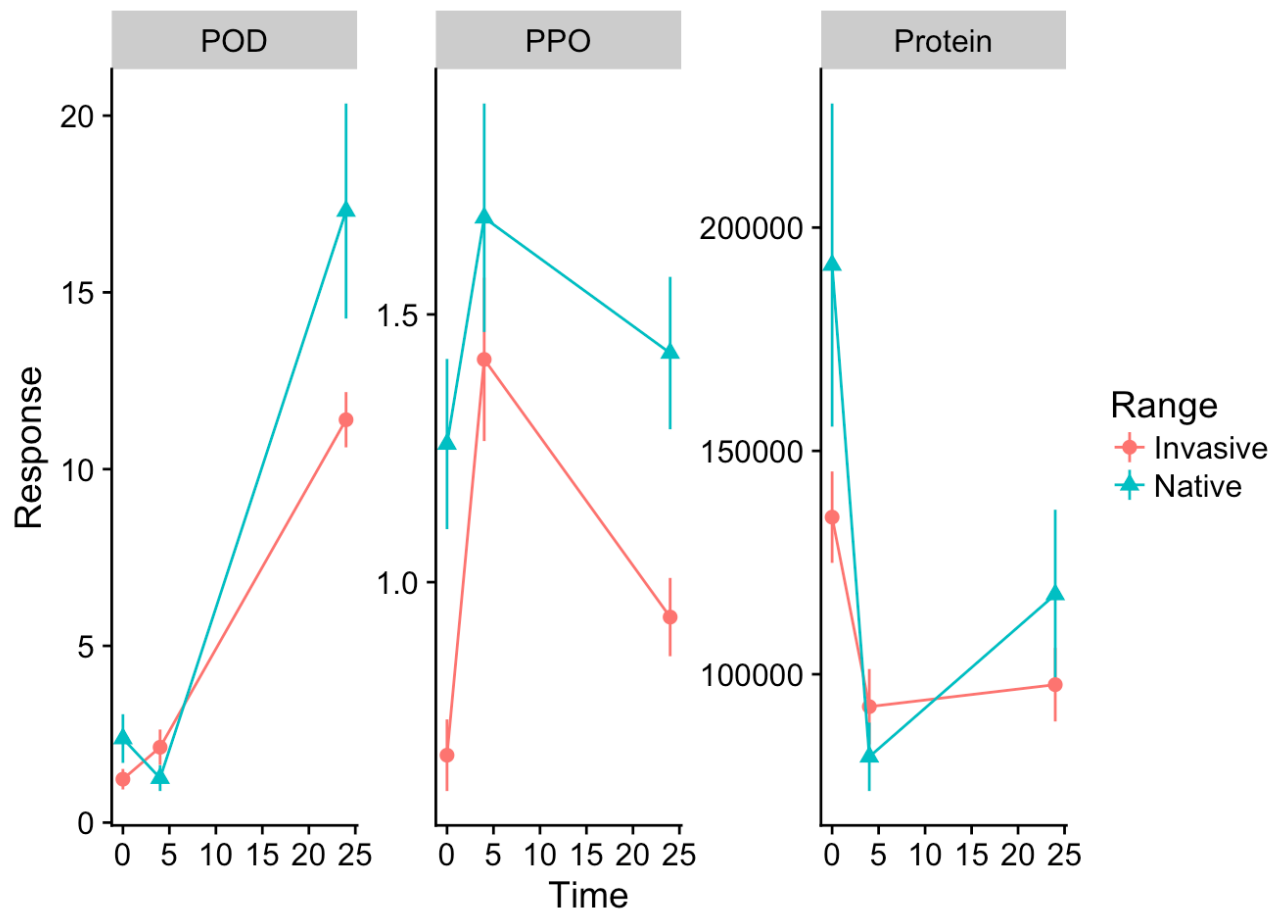
# Add column that identifies each assay
ProteinUntra$Assay <- rep("Protein", nrow(ProteinUntra))
PODUntra$Assay <- rep("POD", nrow(PODUntra))
PPOUntra$Assay <- rep("PPO", nrow(PPOUntra))

# Combine datasets
AllBiocUntrans <- rbind(ProteinUntra, PODUntra)
AllBiocUntrans <- rbind(AllBiocUntrans, PPOUntra)

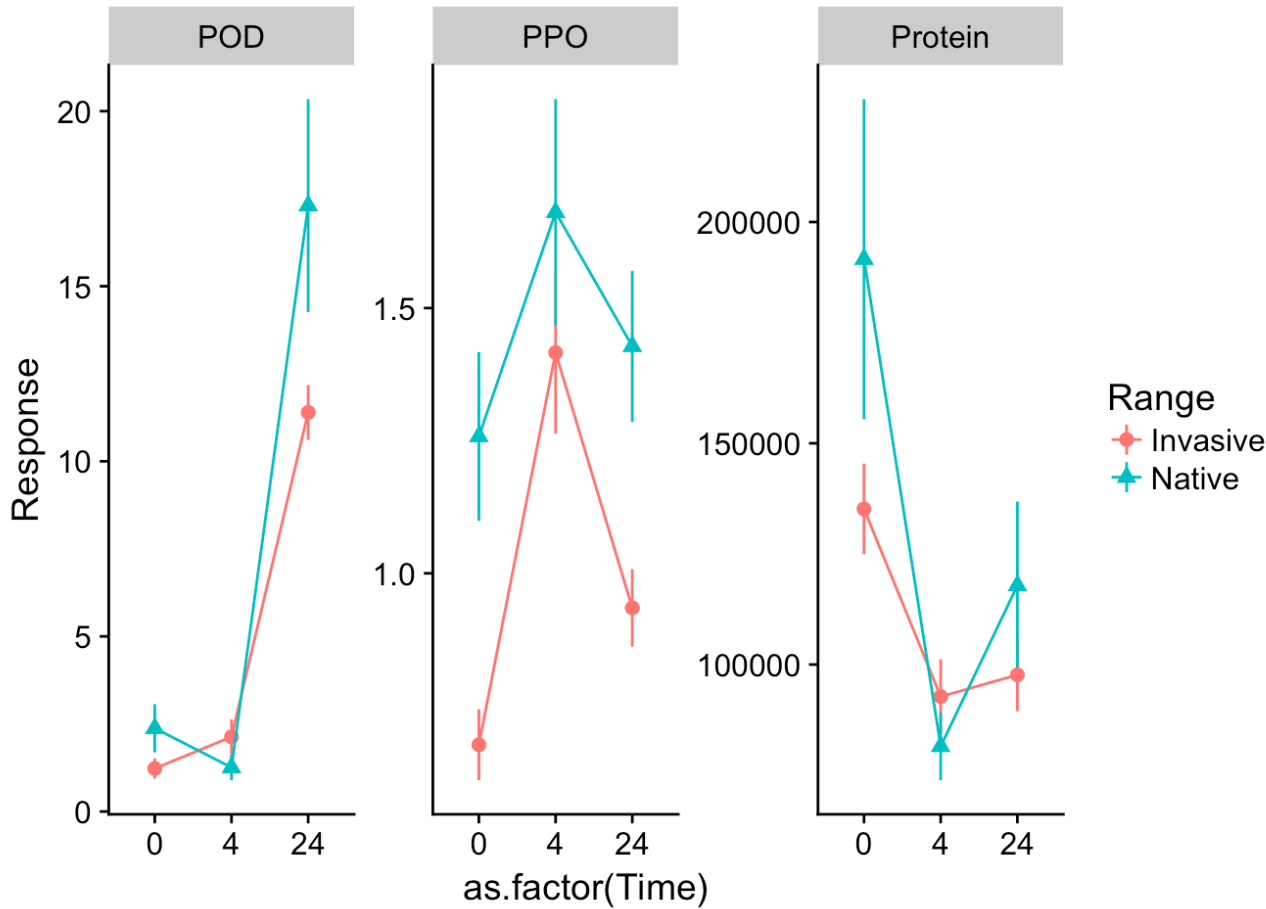
# Make the fig!
ggplot(AllBiocUntrans, aes(x = Time, y = Response, shape = Range, colour = Range)
) + stat_summary(fun.data = "mean_se") + facet_wrap(~Assay, scales = "free") + st
at_summary(fun.y = mean, geom = "line", aes(group = Range))
```

```
## Warning: Removed 12 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 12 rows containing non-finite values (stat_summary).
```



```
ABU <- na.omit(AllBiocUntrans)
ggplot(ABU, aes(x = as.factor(Time), y = Response, shape = Range, colour = Range)) +
  stat_summary(fun.data = "mean_se") + facet_wrap(~Assay, scales = "free") +
  stat_summary(fun.y = mean, geom = "line", aes(group = Range))
```



```

# =====

# ===== Centering MANOVA data Nov 14, 2017=====
center_scale <- function(x) {
  scale(x, scale = FALSE)
}
BiocAssaysWideCenter <- center_scale(BiocAssaysWide[, 6:14])
BiocAssaysWide <- cbind(BiocAssaysWide, BiocAssaysWideCenter)
BAWC <- BiocAssaysWide[, -c(6:14)]

BCmod3 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24, PODHr0, PODHr4, PODHr24, P
POHr0, PPOHr4, PPOHr24) ~ Range, data = BAWC)

Assay <- factor(rep(c("Protein", "POD", "PPO"), each = 3))
MeasTime <- factor(rep(c("Hr0", "Hr4", "Hr24"), 3), levels = c("Hr0", "Hr4", "Hr2
4"))

BCmod4 <- lm(cbind(ProteinHr0, ProteinHr4, ProteinHr24, PODHr0, PODHr4, PODHr24, P
POHr0, PPOHr4, PPOHr24) ~ Site, data = BAWC)

idata <- data.frame(Assay, MeasTime)

anova.BCmod3 <- Anova(BCmod3, idata= idata, idesign = ~ Assay * MeasTime, item =
"Assay:MeasTime", type = 3)

anova.BCmod4 <- Anova(BCmod4, idata= idata, idesign = ~ Assay * MeasTime, item =
"Assay:MeasTime", type = 3)

# ===== Number of each population and by Range Nov 14, 2017 =====
BAWC %>%
  group_by(Site) %>%
  summarise(no_rows = length(Site))

```

```
## # A tibble: 18 x 2
##       Site no_rows
##   <fctr>   <int>
## 1   ARC       18
## 2   AZN        3
## 3   BHI        6
## 4   CAM        6
## 5   CDC       27
## 6   CRG        6
## 7   DCR        3
## 8   DEB       27
## 9   FUS        3
## 10  HIG       30
## 11  HUE        3
## 12  LDS       18
## 13  LIS       15
## 14  MEL        3
## 15  MLR        3
## 16  MUS       21
## 17  NAR       24
## 18  SAN       27
```

```
BAWC %>% count(Range, Site)
```

```
## # A tibble: 18 x 3
##       Range Site      n
##   <fctr> <fctr> <int>
## 1 Invasive ARC       18
## 2 Invasive BHI        6
## 3 Invasive DCR        3
## 4 Invasive MEL        3
## 5 Invasive MLR        3
## 6 Native   AZN        3
## 7 Native   CAM        6
## 8 Native   CDC       27
## 9 Native   CRG        6
## 10 Native  DEB       27
## 11 Native  FUS        3
## 12 Native  HIG       30
## 13 Native  HUE        3
## 14 Native  LDS       18
## 15 Native  LIS       15
## 16 Native  MUS       21
## 17 Native  NAR       24
## 18 Native  SAN       27
```



```
PopulationTable <- BAWC %>% count(Range, Site)
PopulationTable$NoGenotypes <- PopulationTable$n/3
PopTab <- dcast(PopulationTable, Site ~ Range, value.var = "NoGenotypes")

ggplot(PopulationTable, aes(Site, n/3, fill = Range, label = n/3)) + geom_bar(stat = "identity") + ylab("Number of Genotypes") + theme(axis.text.x = element_text(angle = 45)) + geom_text()
```

