Nanosight Analysis

Sean Nguyen

Overall Objective

Load Libraries

```r
library(tidyverse)
library(cowplot)
library(broom)
library(pwr)
```

Import data

```r
setwd("~/GitHub/time-course/data")

rawdata <- "revised_MASTER-ExperimentSummary.csv"
timecourse <- "timecourse2017.csv"


data <- read_csv(rawdata)
tc <- read_csv(timecourse, na = c("","NA"))
```

Convert data from 'wide' to 'long' format

```r
data1 <- data %>%
  gather(Sample,Count,2:226)

# Separate samples by identifiers
data2 <- data1 %>%
  separate(Sample, into=c("Sample_ID","Dilution_factor","Injection","Tech_rep", sep = "_")) %>%
  select(-`_`)
```

Factor the data into categorical variables

```r
# Refactoring Columns for samples
data2$Sample_ID <- as.factor(data2$Sample_ID)
data2$Dilution_factor <- as.numeric(data2$Dilution_factor)
data2$Injection<- as.factor(data2$Injection)
data2$Tech_rep <- as.numeric(data2$Tech_rep)


# Refactoring COlumns for timecourse
tc$Sample_ID <- as.factor(tc$Sample_ID)
tc$Day <- as.factor(tc$Day)
```

```
tc$Weight <- as.numeric(tc$Weight)
tc$TEI_Day <- as.factor(tc$TEI_Day)
```

```
tc1 <- tc %>%
  select(Day:Pups)
tc1
```

```
## # A tibble: 36 × 5
##       Day Sample_ID TEI_Day Weight  Pups
##    <fctr>    <fctr>  <fctr>  <dbl> <int>
## 1       1         1       6  19.60     0
## 2       1         2       2  18.60     0
## 3       1         3       1  19.30     0
## 4       1         4       3  20.50     0
## 5       1         5       4  18.74     0
## 6       1         6       5  18.30     0
## 7       5         7       6  20.83    10
## 8       5         8       3  19.88    10
## 9       5         9       2  24.41     9
## 10      5        10       4  20.01     6
## # ... with 26 more rows
```

Back calculate the original concentration of the sample

```
data2 <- data2 %>%
  mutate(True_Count=Dilution_factor*Count)
data2
```

```
## # A tibble: 225,000 × 7
##     particle_size Sample_ID Dilution_factor Injection Tech_rep Count
##             <dbl>    <fctr>           <dbl>    <fctr>    <dbl> <int>
## 1             0.5         2             500         1        0     0
## 2             1.5         2             500         1        0     0
## 3             2.5         2             500         1        0     0
## 4             3.5         2             500         1        0     0
## 5             4.5         2             500         1        0     0
## 6             5.5         2             500         1        0     0
## 7             6.5         2             500         1        0     0
## 8             7.5         2             500         1        0     0
## 9             8.5         2             500         1        0     0
## 10            9.5         2             500         1        0     0
## # ... with 224,990 more rows, and 1 more variables: True_Count <dbl>
```

Average the three technical replicate readings

```
data3 <- data2 %>%
  group_by(particle_size,Sample_ID,Dilution_factor,Injection) %>%
  summarise( tech_N = length(True_Count),
             tech_mean = mean(True_Count),
             tech_sd = sd(True_Count),
             tech_se = tech_sd/sqrt(tech_N))
data3
```

```
## Source: local data frame [74,000 x 8]
## Groups: particle_size, Sample_ID, Dilution_factor [?]
##
##    particle_size Sample_ID Dilution_factor Injection tech_N tech_mean
##            <dbl>    <fctr>           <dbl>    <fctr>  <int>     <dbl>
## 1            0.5         1             500         1      3         0
## 2            0.5         1             500         2      3         0
## 3            0.5        10             500         1      3         0
## 4            0.5        10             500         2      3         0
## 5            0.5        11             500         1      3         0
## 6            0.5        11             500         2      3         0
## 7            0.5        12             500         1      6         0
## 8            0.5        12             500         2      3         0
## 9            0.5        13             500         1      3         0
## 10           0.5        13             500         2      3         0
## # ... with 73,990 more rows, and 2 more variables: tech_sd <dbl>,
## #   tech_se <dbl>
```

```r
test1 <- left_join(tc1,data3, by= "Sample_ID")
```

Summarize samples by injection (average both injections)

```r
data4 <- data3 %>%
  group_by(particle_size,Sample_ID,Dilution_factor) %>%
  summarise( inj_N = length(tech_mean),
            inj_mean = mean(tech_mean),
            inj_sd = sd(tech_mean),
            inj_se = inj_sd/sqrt(inj_N))
data4
```

```
## Source: local data frame [37,000 x 7]
## Groups: particle_size, Sample_ID [?]
##
##    particle_size Sample_ID Dilution_factor inj_N inj_mean inj_sd inj_se
##            <dbl>    <fctr>           <dbl> <int>    <dbl>  <dbl>  <dbl>
## 1            0.5         1             500     2        0      0      0
## 2            0.5        10             500     2        0      0      0
## 3            0.5        11             500     2        0      0      0
## 4            0.5        12             500     2        0      0      0
## 5            0.5        13             500     2        0      0      0
## 6            0.5        14             500     2        0      0      0
## 7            0.5        15             500     2        0      0      0
## 8            0.5        16             500     2        0      0      0
## 9            0.5        17             500     2        0      0      0
## 10           0.5        18             500     2        0      0      0
## # ... with 36,990 more rows
```

```r
test2 <- left_join(tc1,data4, by= "Sample_ID")

test2
```

```
## # A tibble: 37,000 × 11
##      Day Sample_ID TEI_Day Weight  Pups particle_size Dilution_factor
##   <fctr>     <chr>  <fctr>  <dbl> <int>         <dbl>           <dbl>
```

```
## 1        1        1        6    19.6    0            0.5              500
## 2        1        1        6    19.6    0            1.5              500
## 3        1        1        6    19.6    0            2.5              500
## 4        1        1        6    19.6    0            3.5              500
## 5        1        1        6    19.6    0            4.5              500
## 6        1        1        6    19.6    0            5.5              500
## 7        1        1        6    19.6    0            6.5              500
## 8        1        1        6    19.6    0            7.5              500
## 9        1        1        6    19.6    0            8.5              500
## 10       1        1        6    19.6    0            9.5              500
## # ... with 36,990 more rows, and 4 more variables: inj_N <int>,
## #   inj_mean <dbl>, inj_sd <dbl>, inj_se <dbl>
```

Quick visualizations

Graphing all samples

```r
test1$Sample_ID_correct = factor(test1$Sample_ID, levels=c('1','2','3','4','5','6','7','8','9','10','11

graph1 <- test1 %>%
  ggplot(aes(x=particle_size, y=tech_mean,color=Injection ))+ #plot
  geom_ribbon(aes(ymin=tech_mean-tech_se, ymax=tech_mean+tech_se),alpha=0.2,fill = alpha('grey12', 0.2)
  geom_line(size=2.0) + xlim(0,500)+ #line size, x-axis scale
  scale_y_continuous(expand=c(0,0))+ #set bottom of graph
  xlab("Particle Size") + # X axis label
  ylab("\nMean Particle Concentration/ml\n") + # Y axis label
  ggtitle("Nanosight Histogram of\nVirgin Mouse Plasma")+ #title
  labs(color="Injection")+ #Label table title
  facet_wrap( ~ Sample_ID_correct, nrow=6)

graph1
```

Nanosight Histogram of
Virgin Mouse Plasma
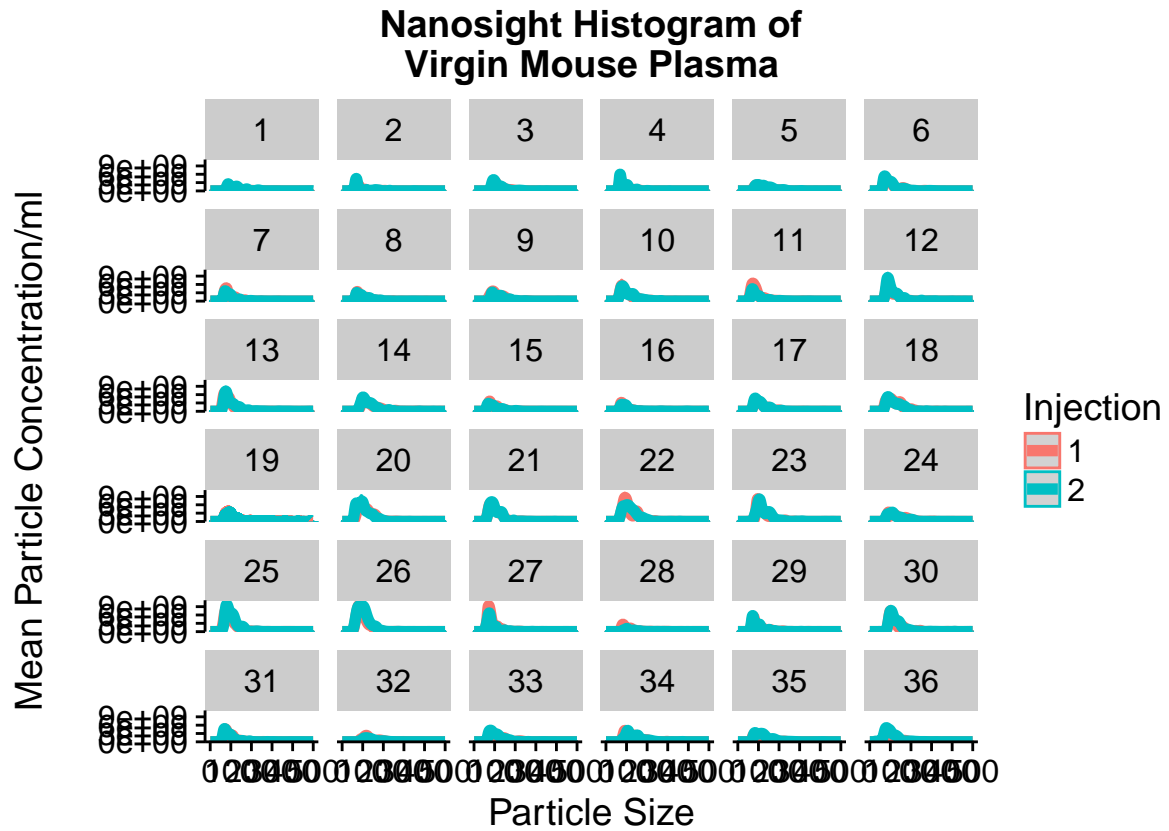
Graphing averaged samples by experimental day

```
graph2 <- test2 %>%
  group_by(TEI_Day) %>%
  ggplot(aes(x=particle_size, y=inj_mean,color=Day ))+ #plot
  #geom_ribbon(aes(ymin=inj_mean-inj_se, ymax=inj_mean+inj_se),alpha=0.2,fill = alpha('grey12', 0.2)) +
  geom_line(size=2) + xlim(0,500)+ #line size, x-axis scale
  scale_y_continuous(expand=c(0,0))+ #set bottom of graph
  xlab("Particle Size") + # X axis label
  ylab("\nMean Particle Concentration/ml\n") + # Y axis label
  ggtitle("Nanosight Histogram of\nVirgin Mouse Plasma")+ #title
  labs(color="Condition")+ #Label table title
  facet_wrap(~ TEI_Day, ncol=6)

graph2
```

Nanosight Histogram of
Virgin Mouse Plasma
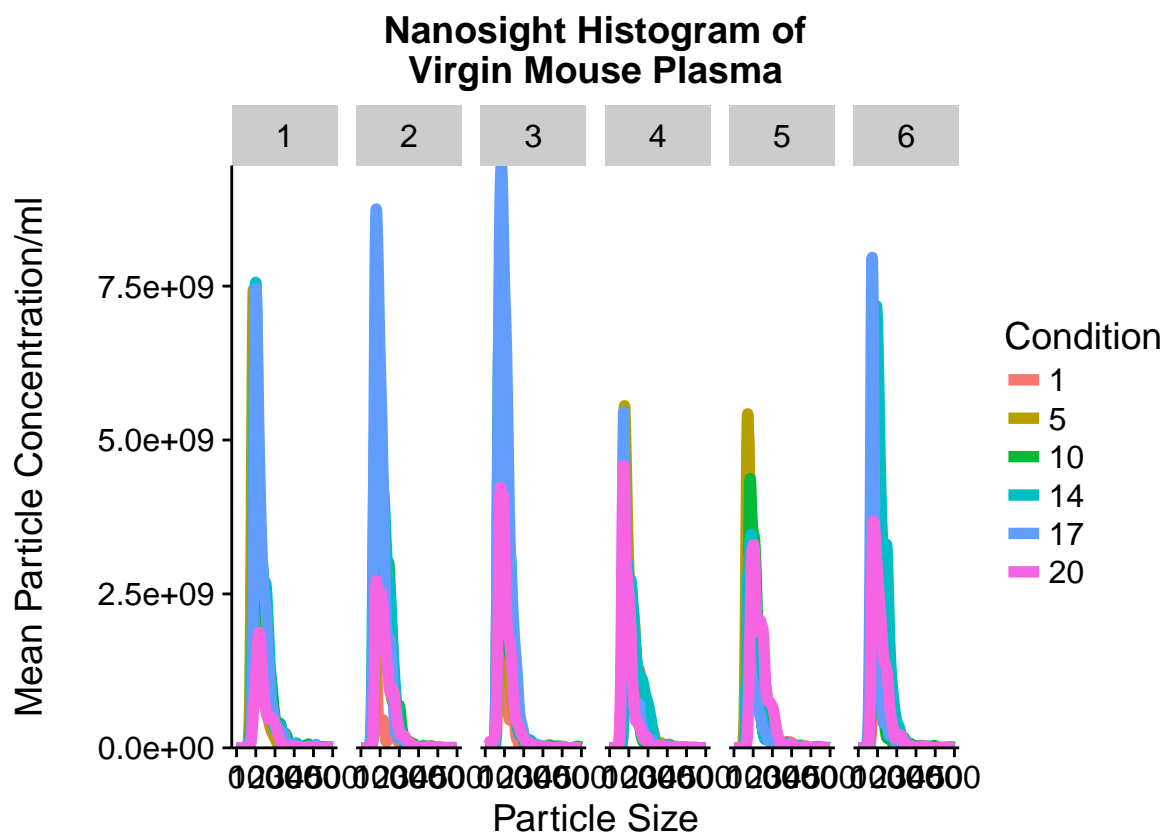
Particle concentration values for each of the 36 samples

```
test3 <- test2 %>%
  group_by(Day,Sample_ID) %>%
  summarise(particle_conc=sum(inj_mean))
test3
```

```
## Source: local data frame [36 x 3]
## Groups: Day [?]
##
##       Day Sample_ID particle_conc
##    <fctr>     <chr>         <dbl>
## 1       1         1  124235076333
## 2       1         2   89249062500
## 3       1         3  175595167167
## 4       1         4  153741474667
## 5       1         5  182057049833
## 6       1         6  282458569250
## 7       5        10  306075805167
## 8       5        11  241497990000
## 9       5        12  336677212208
## 10      5         7  167230730583
## # ... with 26 more rows
```

Summary statistics of particle concentration (averaging n=6 for each time point)

```
test4 <- test3 %>%
  group_by(Day) %>%
  summarise(Day_N=length(particle_conc),
            Day_mean = mean(particle_conc),
            Day_sd = sd(particle_conc),
            Day_se = Day_sd/sqrt(Day_N))
```

Boxplot

```
plot1 <- test3 %>%
  group_by(Day) %>%
  ggplot(aes(factor(Day),particle_conc, color=Day)) +
  geom_boxplot(colour="black",fill=NA) +
  geom_point(position='jitter',size=3)+
  xlab("\nDay of Gestation\n") + # X axis label
  ylab("\nExosomes/ml\n") + # Y axis label
  ggtitle("Plasma Exosome Concentration\nThroughout Pregnancy\n")+ #title
  labs(color="Condition")#Label table title
plot1
```
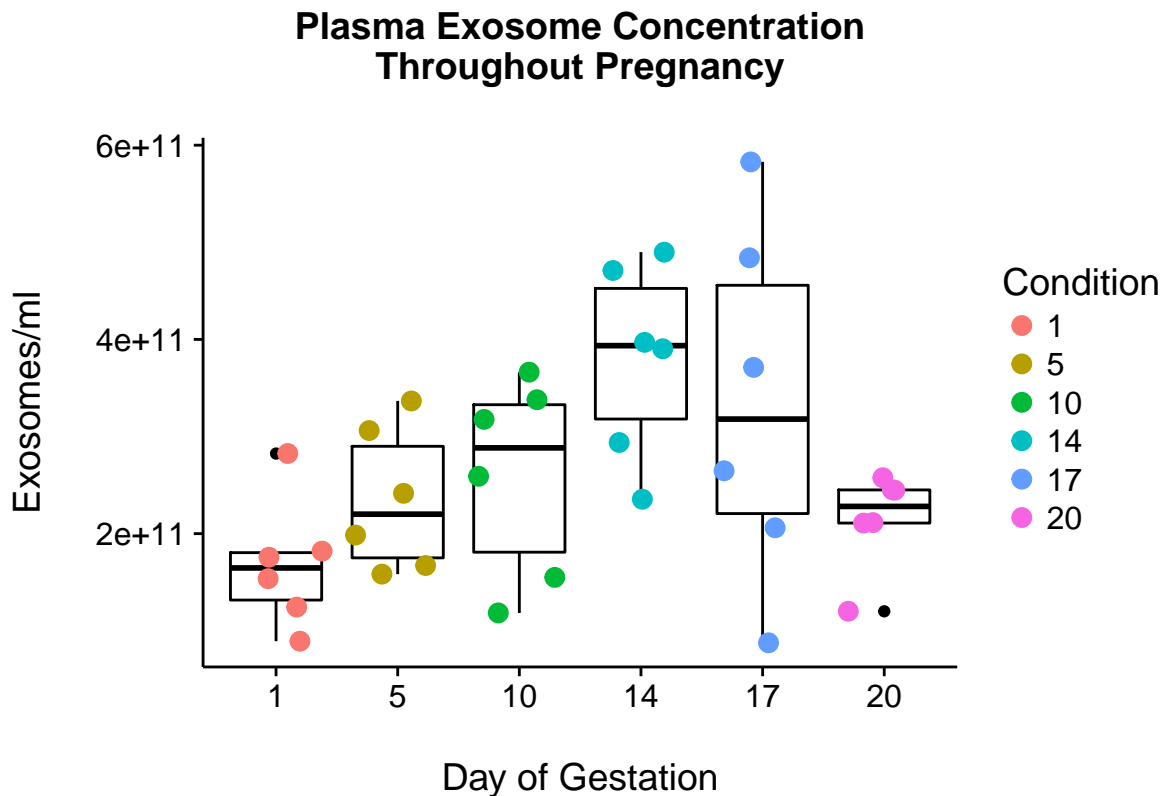
Bar plot

```
plot <- test4 %>%
  ggplot(aes(x=Day, y=Day_mean, fill=Day ))+ #plot
  geom_col()+
  geom_errorbar(aes(ymin=Day_mean-Day_se, ymax=Day_mean+Day_se), width=.5,
                size=0.8, colour="black", position=position_dodge(.9)) + #error bars
  scale_y_continuous(expand=c(0,0))+ #set bottom of graph
  xlab("\nDay of Gestation\n") + # X axis label
  ylab("\nExosomes/ml\n") + # Y axis label
  ggtitle("Plasma Exosome Concentration\nThroughout Pregnancy\n")+ #title
  labs(color="Condition")#Label table title
plot
```
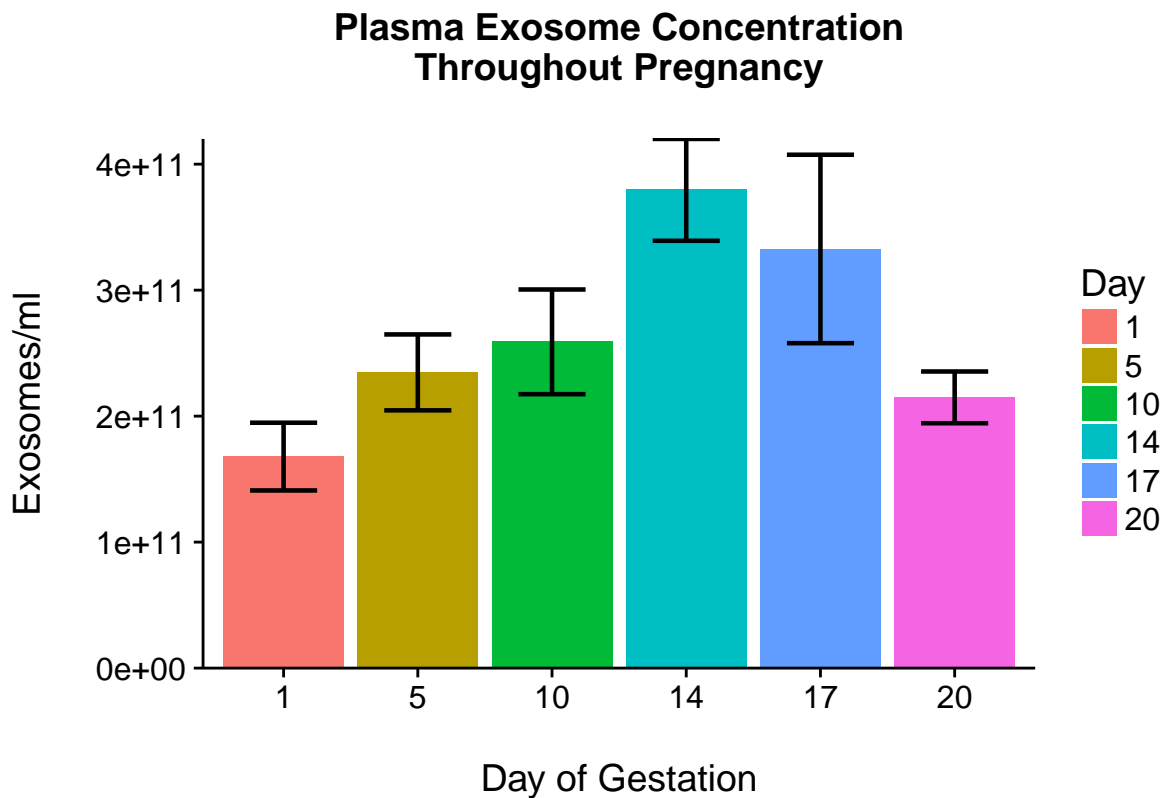

Looking at Variation between the days the samples were run

```
test7 <- test3 %>%
  left_join(tc1)
```

```
## Joining, by = c("Day", "Sample_ID")
```

```
plot2 <- test7 %>%
  ggplot(aes(x = Day, y = particle_conc, color = Day, shape=TEI_Day))+
  geom_point(position= 'dodge',size=4)+
  scale_shape_manual(values=c(15,16,17,22,23,24))+
```
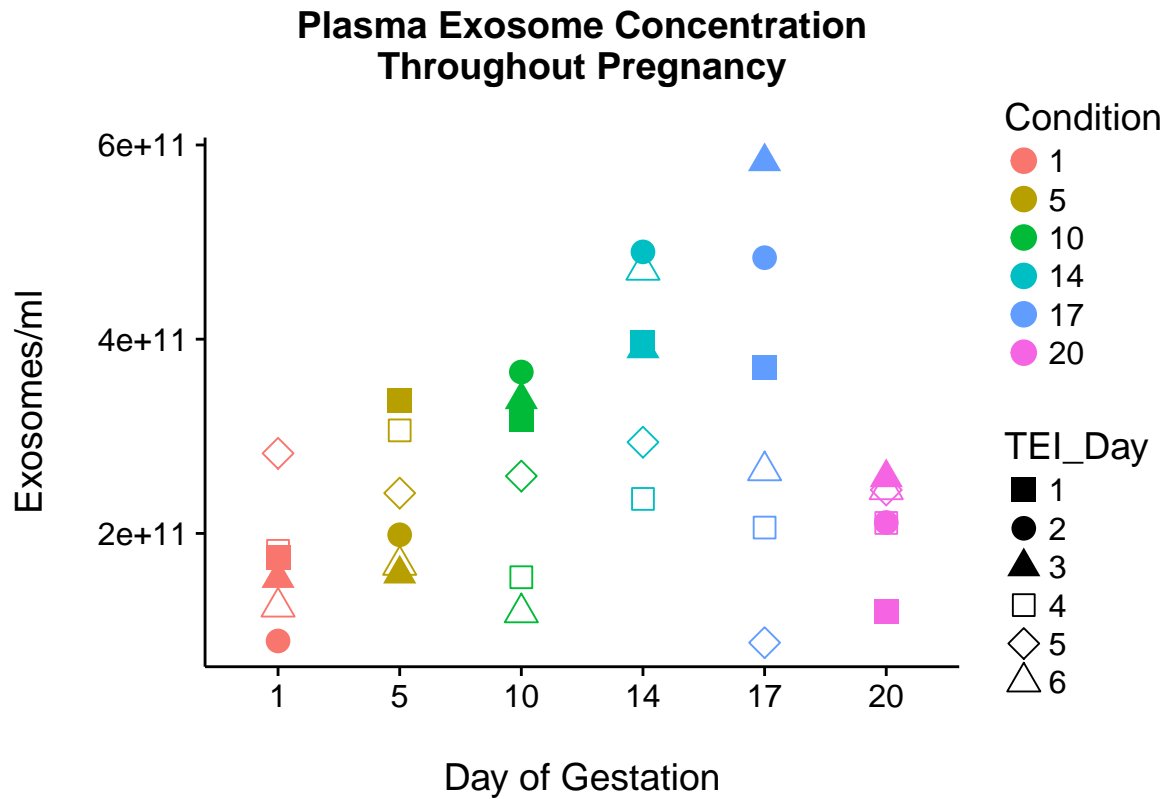
```
    xlab("\nDay of Gestation\n") + # X axis label
    ylab("\nExosomes/ml\n") + # Y axis label
    ggtitle("Plasma Exosome Concentration\nThroughout Pregnancy\n")+ #title
    labs(color="Condition")#Label table title
```

```
plot2
```

**Plasma Exosome Concentration
Throughout Pregnancy**

Statistics

Shapiro test

```
tidy(shapiro.test(test3$particle_conc))
```

```
##   statistic  p.value                    method
## 1 0.9513772 0.1155555 Shapiro-Wilk normality test
```

p value >0.05 therefore conlude data is normally distributed

ANOVA

```
fit <- aov(particle_conc ~ Day, data=test3)
stats <- tidy(fit)
stats
```

```
##          term df        sumsq        meansq statistic    p.value
## 1        Day  5 1.836522e+23 3.673044e+22  3.337966 0.01627918
## 2 Residuals 30 3.301152e+23 1.100384e+22        NA         NA
```

Statistically significant, thus Tukey's HSD post hoc analysis can determine significant differences.

Tukey Post Hoc Test

```
HSD <- TukeyHSD(fit)
tukey <- tidy(HSD)
tukey
```

```
##      term comparison       estimate       conf.low     conf.high adj.p.value
## 1    Day        5-1    66832709979 -117377107609  251042527567   0.8760587
## 2    Day       10-1    91084735500  -93125082088  275294553088   0.6644418
## 3    Day       14-1   211706689372   27496871784  395916506960   0.0169879
## 4    Day       17-1   164807106233  -19402711355  349016923822   0.1003839
## 5    Day       20-1    47007296097 -137202521491  231217113685   0.9694889
## 6    Day       10-5    24252025521 -159957792067  208461843109   0.9985338
## 7    Day       14-5   144873979393  -39335838195  329083796981   0.1911489
## 8    Day       17-5    97974396254  -86235421334  282184213842   0.5937845
## 9    Day       20-5   -19825413882 -204035231470  164384403706   0.9994463
## 10   Day      14-10   120621953872  -63587863716  304831771460   0.3704012
## 11   Day      17-10    73722370733 -110487446855  257932188322   0.8250404
## 12   Day      20-10   -44077439403 -228287256991  140132378185   0.9768879
## 13   Day      17-14   -46899583139 -231109400727  137310234449   0.9697867
## 14   Day      20-14  -164699393275 -348909210863   19510424313   0.1007550
## 15   Day      20-17  -117799810136 -302009627724   66410007452   0.3960640
```

Significant Tukey Post Hoc Test Values

```
tukey %>%
  filter(adj.p.value<0.05) %>%
  arrange(adj.p.value)
```

```
##   term comparison      estimate    conf.low    conf.high adj.p.value
## 1  Day       14-1 211706689372 27496871784 395916506960   0.0169879
```