tidy\_field\_horses\_raw\_counts <- field\_horses\_raw\_counts\_modified %>%

gather(key = horse\_id, value = counts, -gene)

tidy\_field\_horses\_raw\_counts

full\_data <- full\_join(tidy\_field\_horses\_raw\_counts, field\_horses\_metadata\_modified)

full\_data

write.csv(full\_data,"results/field\_horses\_full\_data.csv")

# raw counts of microRNA of 6 horses with geom\_plot

full\_data\_plot1 <- ggplot (data = full\_data,

aes(x = horse\_id,

y = counts,

color = infection,

group = horse\_id))+

geom\_point()+

scale\_y\_log10()+

scale\_color\_manual(values = c(mock = "grey", infected = "red")) +

theme\_bw() +

labs(title = "microRNA raw counts of horses together")

full\_data\_plot1

ggsave(filename = "results/Figure 1.png", plot = full\_data\_plot1, width = 10, height = 8, dpi = 300, units = "cm")

# raw counts of microRNA of ech horse with geom\_boxplot and geom\_jitter

full\_data\_plot2 <- ggplot (data = full\_data,

aes(x = horse\_id,

y = counts,

color = infection,

group = horse\_id))+

geom\_boxplot() +

geom\_jitter()+

scale\_y\_log10()+

scale\_color\_manual(values = c(mock = "grey", infected = "red")) +

theme\_bw() +

labs(title = "microRNA raw counts of 6 horses")

full\_data\_plot2

ggsave(filename = "results/Figure 2.png", plot = full\_data\_plot2, width = 12, height = 8, dpi = 300, units = "cm")

# plot figure with raw count > 1000 for all horses with geom\_boxplot() + geom\_jitter()

raw\_count\_more\_than\_1000 <- full\_data %>%

filter(counts > 1000)

raw\_count\_more\_than\_1000

plot3 <- ggplot(data = raw\_count\_more\_than\_1000,

aes(x = horse\_id,

y = counts,

color = infection,

group = horse\_id))+

geom\_boxplot() +

geom\_jitter()+

scale\_y\_log10()+

scale\_color\_manual(values = c(mock = "grey", infected = "red")) +

theme\_bw() +

labs(title = "microRNA raw counts more than 1000 of 6 horses")

plot3

ggsave(filename = "results/Figure 3.png", plot = plot3, width = 12, height = 8, dpi = 300, units = "cm")

# plot figure with raw count < 1000 for all horses with geom\_boxplot() + geom\_jitter()

raw\_count2 <- full\_data %>%

filter(counts < 1000)

raw\_count2

plot4 <- ggplot(data = raw\_count2,

aes(x = horse\_id,

y = counts,

color = infection,

group = horse\_id))+

geom\_boxplot() +

geom\_jitter()+

scale\_y\_log10()+

scale\_color\_manual(values = c(mock = "grey", infected = "red")) +

theme\_bw() +

labs(title = "microRNA raw counts < 1000 of 6 horses")

plot4

ggsave(filename = "results/Figure 4.png", plot = plot4, width = 12, height = 8, dpi = 300, units = "cm")

# Do some PCA analysis with full\_data

full\_data

new\_full\_data <- full\_data %>%

select(-horse\_id, -infection) %>%

mutate(gene = as.numeric(gene))

new\_full\_data

mutate(horse\_id = as.numeric(horse\_id), infection = as.numeric(horse\_id) )

new\_full\_data\_scaled <- select(full\_data, -gene, -horse\_id, -infection) %>% scale

head(new\_full\_data\_scaled)

new\_full\_data\_pca <- prcomp(new\_full\_data\_scaled)

summary(new\_full\_data\_pca)

plot(new\_full\_data\_pca)