4.Gene_expressions_analysis - Analyzing differences between field and lab

Fay Webster

2022-07-18

Aims:

- Compare immune gene expression between lab and field data
- How do eimeria infections affect the expression of various immune genes?
- Does the expression vary between lab and field data?

load libraries

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(optimx)
```

Import data:

Here, we have the experimental / field data, inlcuding imputed data

```
lab <- read.csv("output_data/gene_expression/data_products/lab_imputed_gene_expression.csv")
field <- read.csv("output_data/gene_expression/data_products/field_imputed_gene_expression.csv")
```

Selecting genes

```
# vectors for selecting gene columns
Genes_lab <- c("IFNy", "CXCR3", "IL.6", "IL.10", "IL.13", "IL1RN", "CASP1",
               "CXCL9", "IDO1", "IRGM1", "MPO", "MUC2", "MUC5AC",
           "MYD88", "NCR1", "PRF1", "RETNLB", "SOCS1", "TICAM1", "TNF")

Genes_field   <- c("IFNy",  "CXCR3", "IL.6", #"GBP2", "IL.12", "IRG6",
                   "IL.10", "IL.13", "IL1RN",
                   "CXCR3", "CASP1", "CXCL9",
```
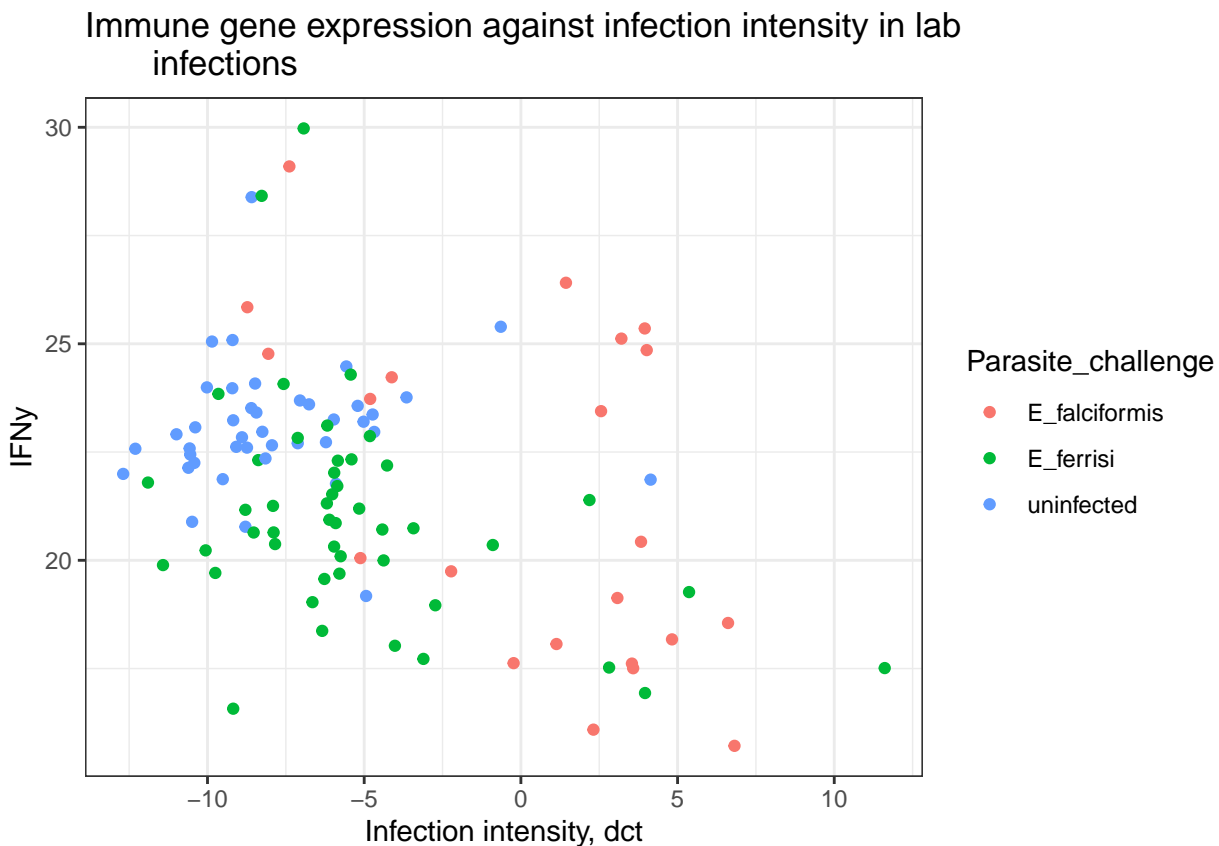
```
                    "IDO1", "IRGM1", "MPO", "MUC2", "MUC5AC", "MYD88",
                    "NCR1", "PRF1", "RETNLB", "SOCS1", "TICAM1", "TNF")
```

1. IFNy

```
ggplot(lab, aes(x = delta, y = IFNy, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```
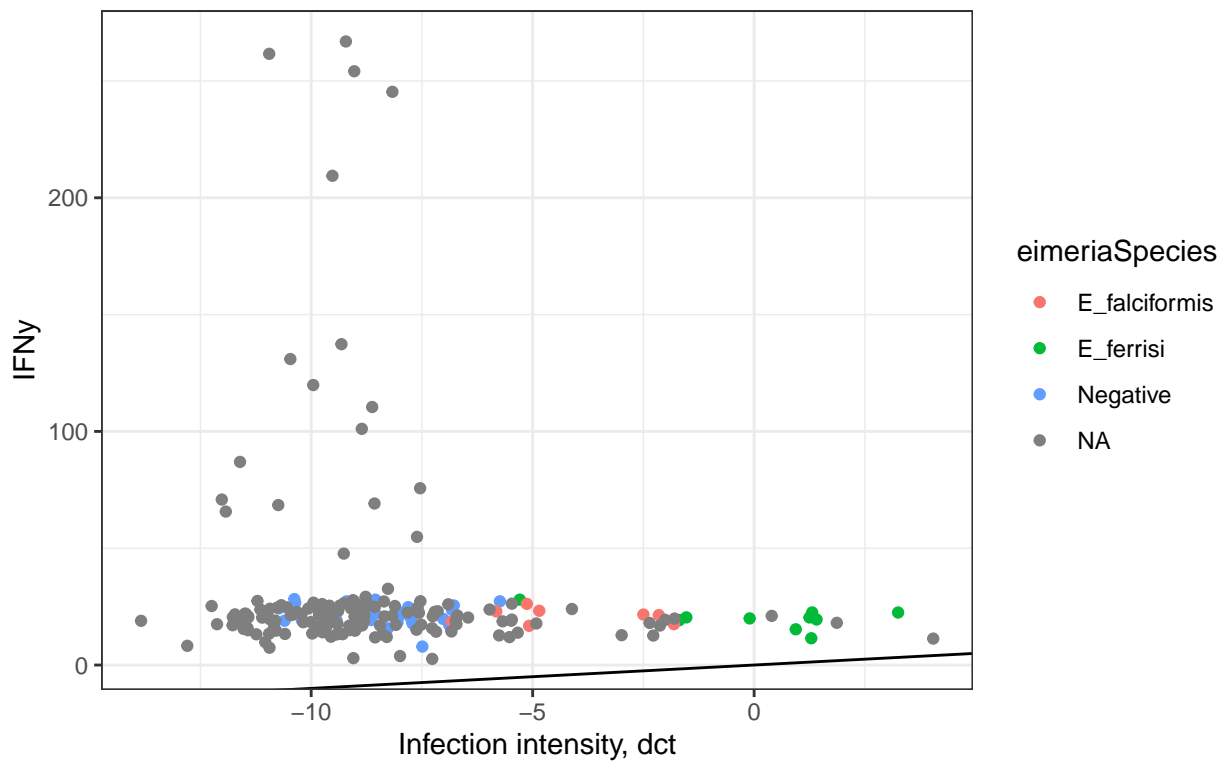
Warning: Removed 6 rows containing missing values (geom_point).


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IFNy, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

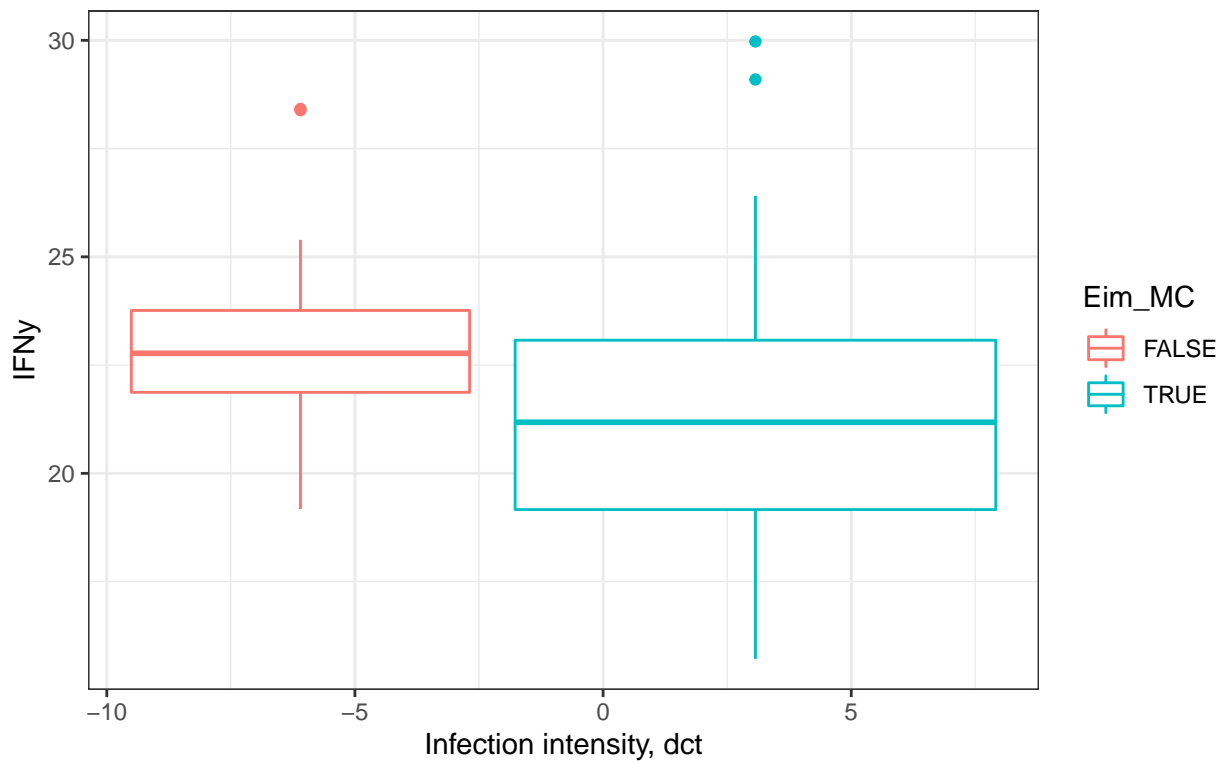## Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples

```
ggplot(lab, aes(x = delta, y = IFNy, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```
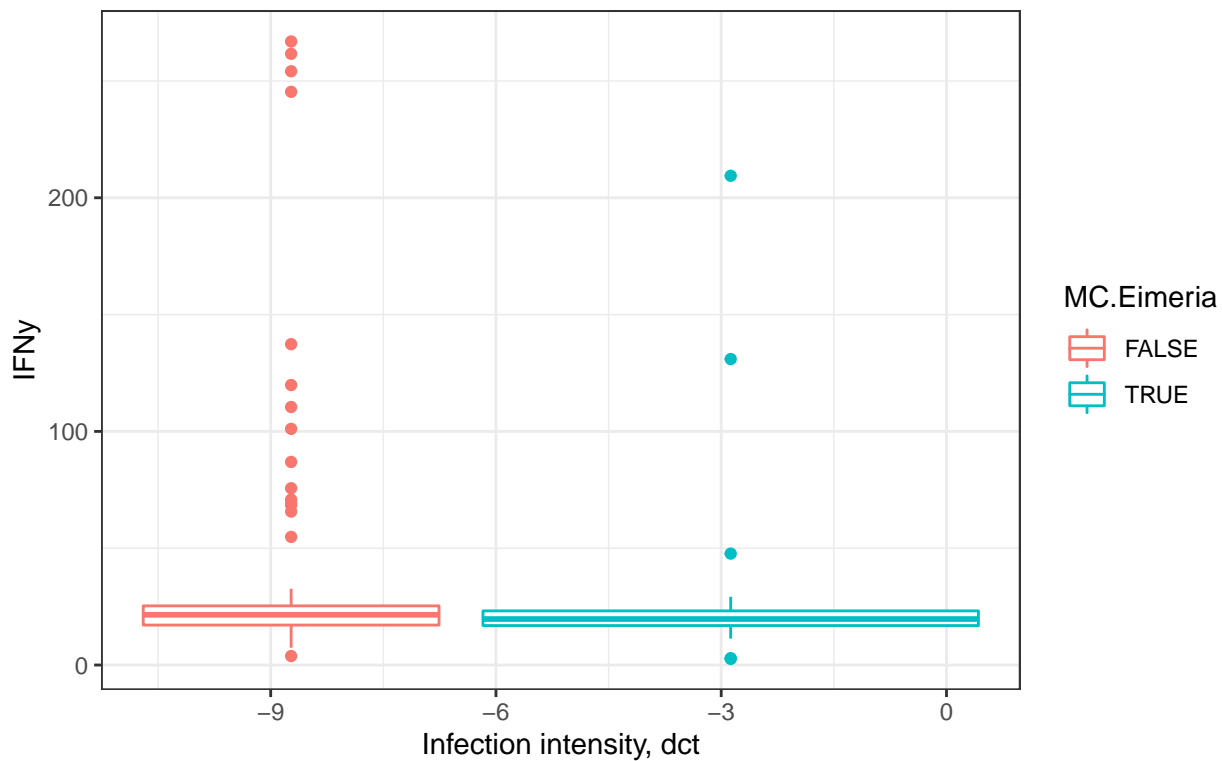
Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IFNy,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

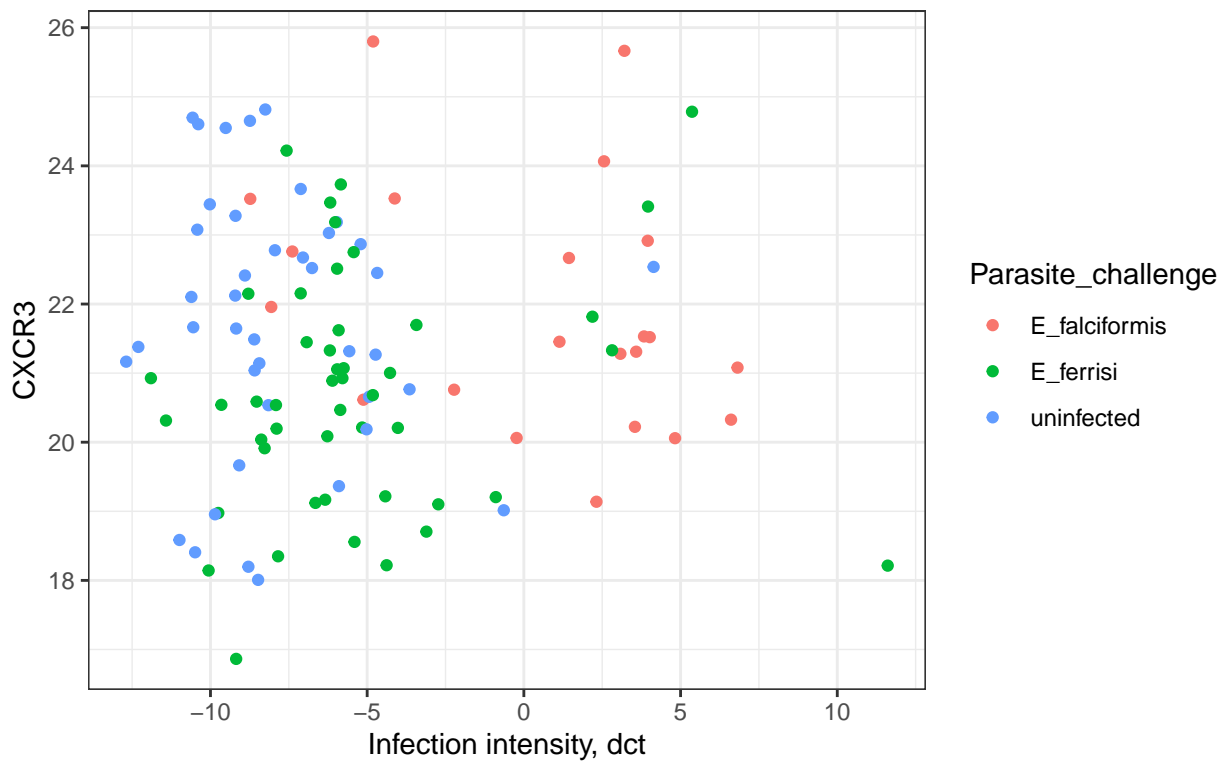Immune gene expression against infection intensity in field samples

2. CXCR3

```
ggplot(lab, aes(x = delta, y = CXCR3_bio, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct", y = "CXCR3",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
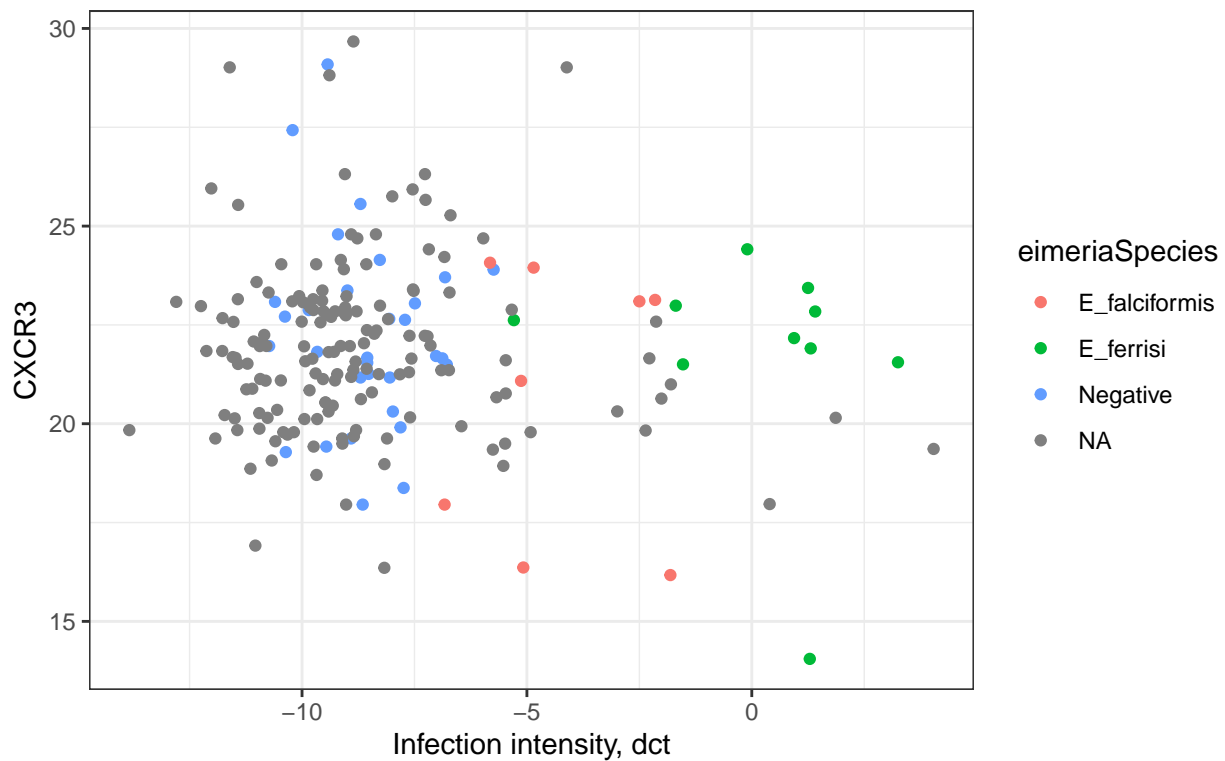
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = CXCR3, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in
       field samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```

Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = CXCR3, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```
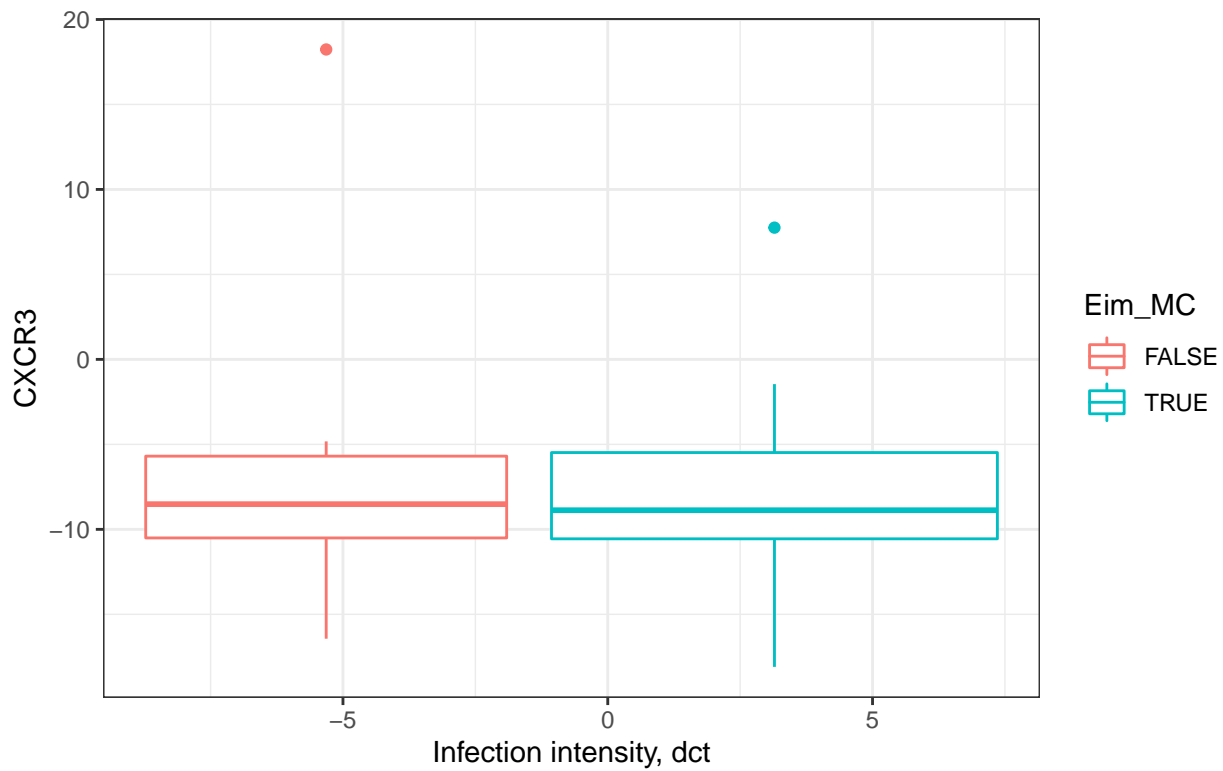
Warning: Removed 6 rows containing missing values (stat_boxplot).

Warning: Removed 49 rows containing non-finite values (stat_boxplot).

Immune gene expression against infection intensity in lab infections
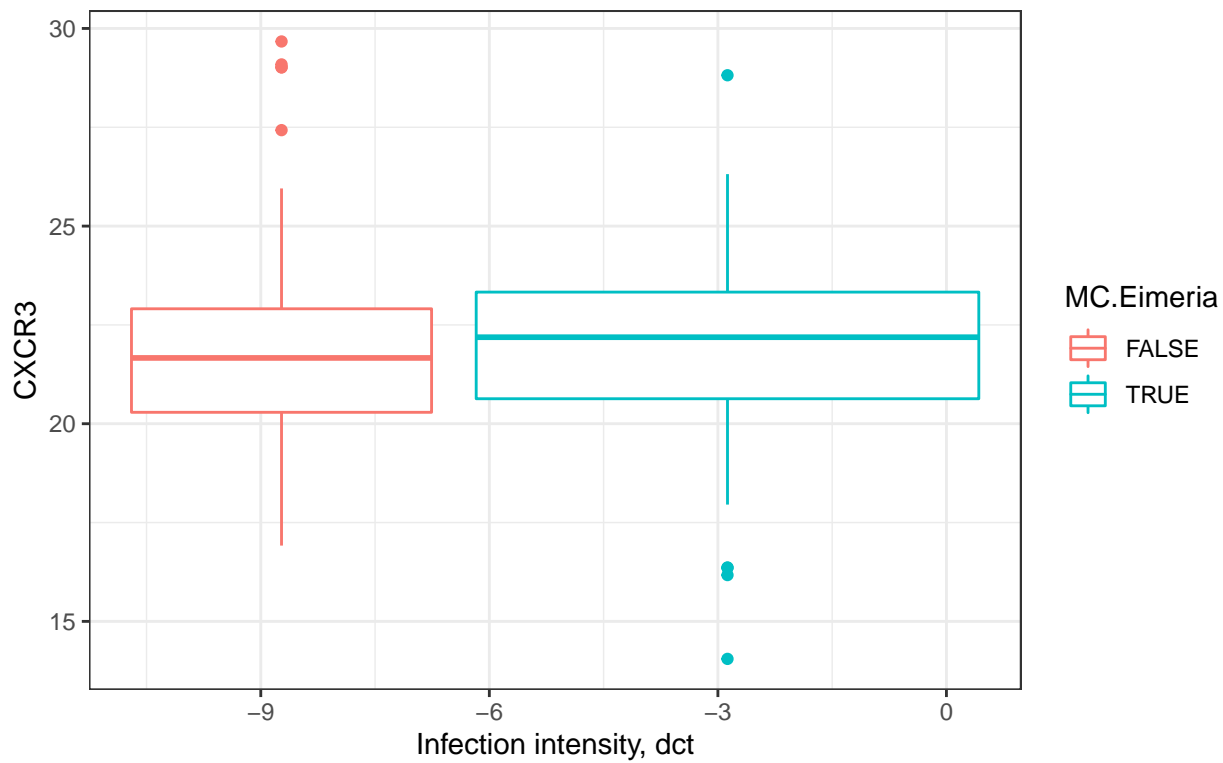


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = CXCR3,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

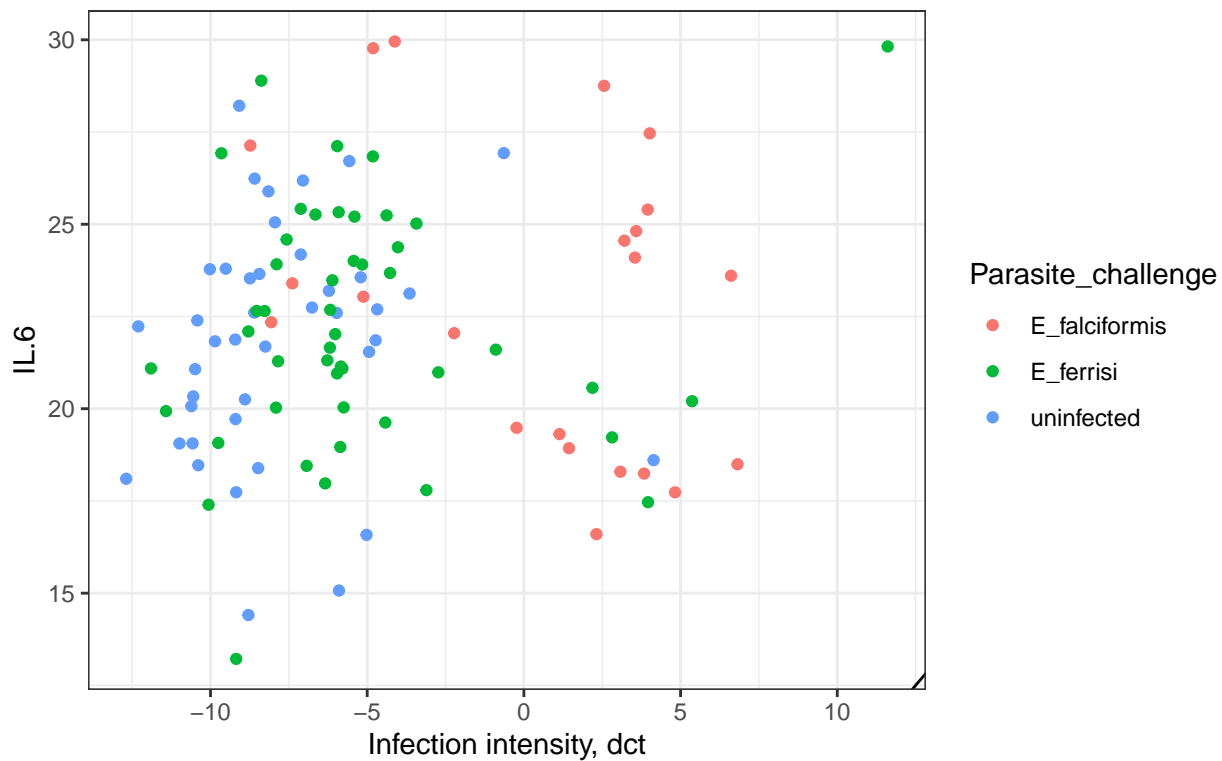Immune gene expression against infection intensity in field samples

3. IL.6

```
ggplot(lab, aes(x = delta, y = IL.6, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
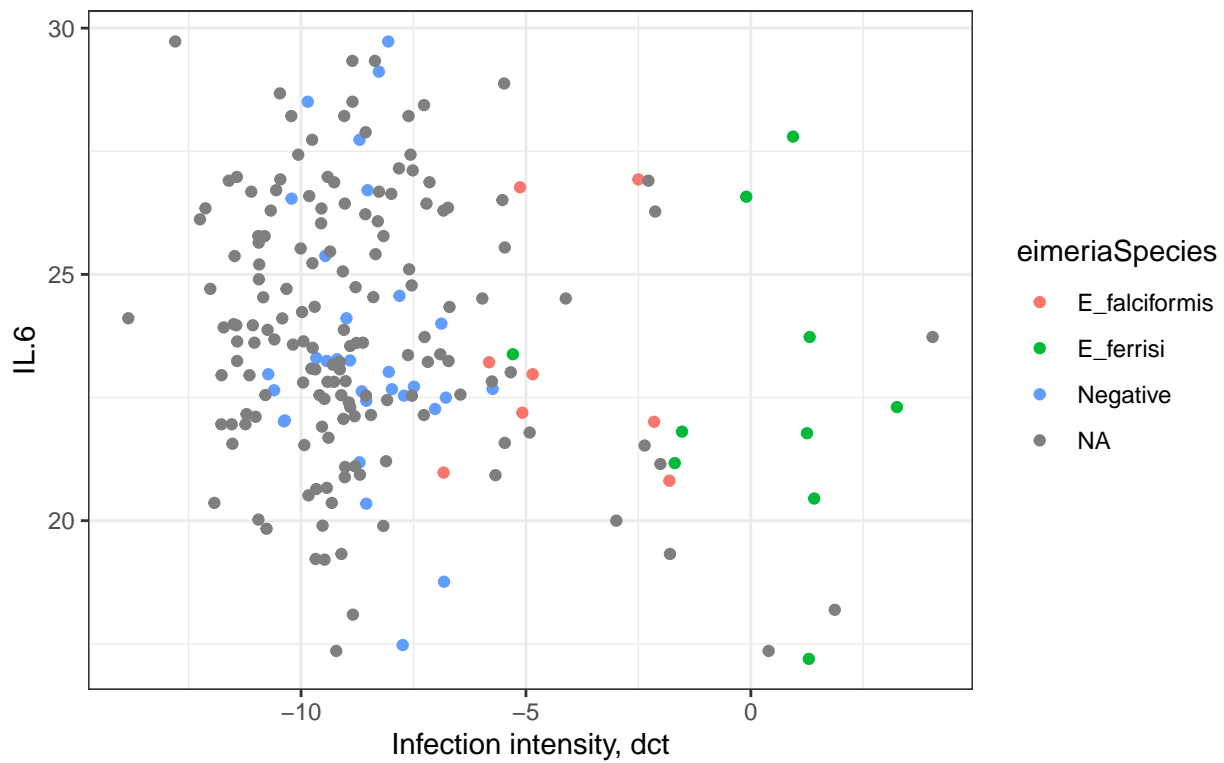
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL.6, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```

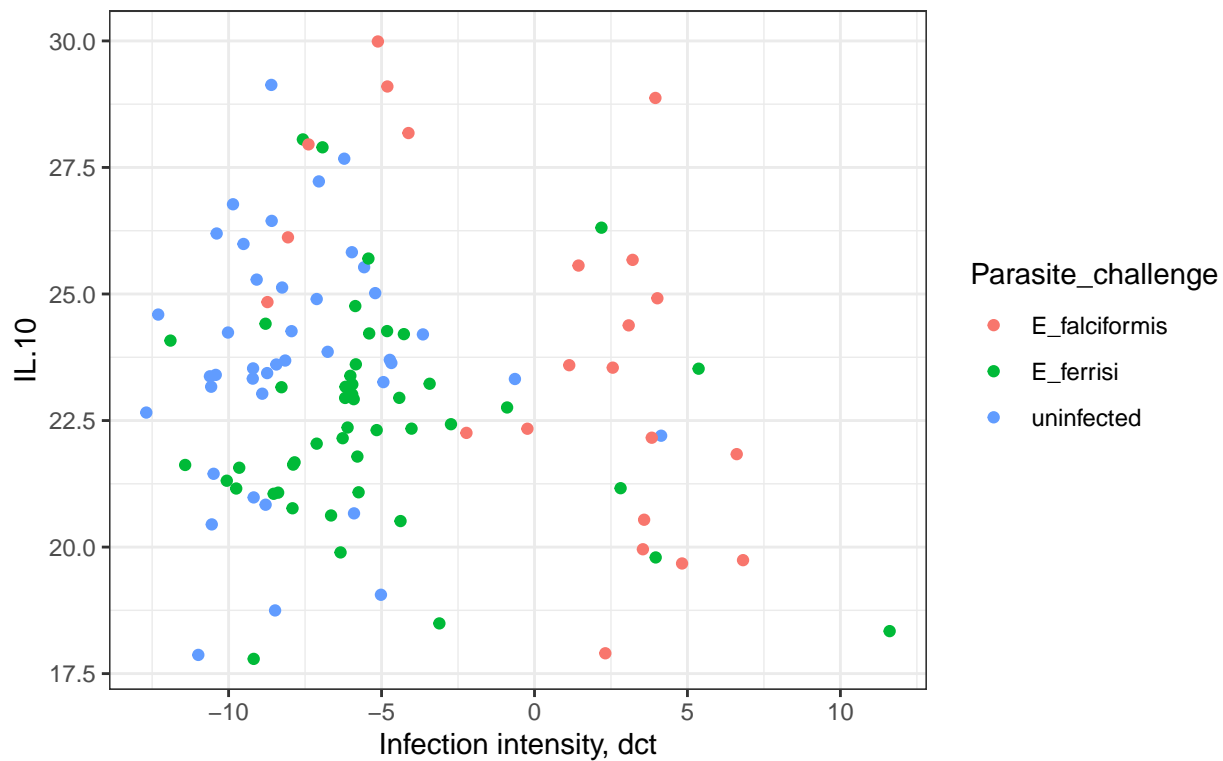Immune gene expression against infection intensity in field samples

4. IL.10

```
ggplot(lab, aes(x = delta, y = IL.10, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
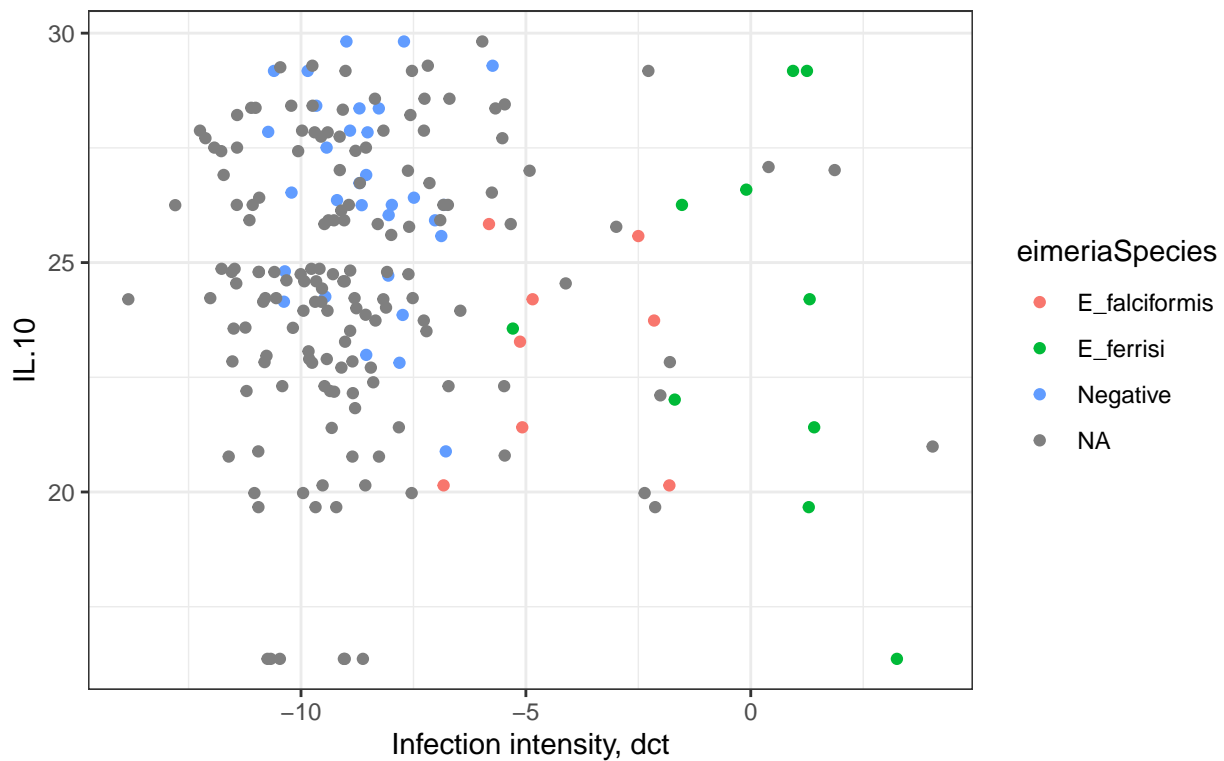
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL.10, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

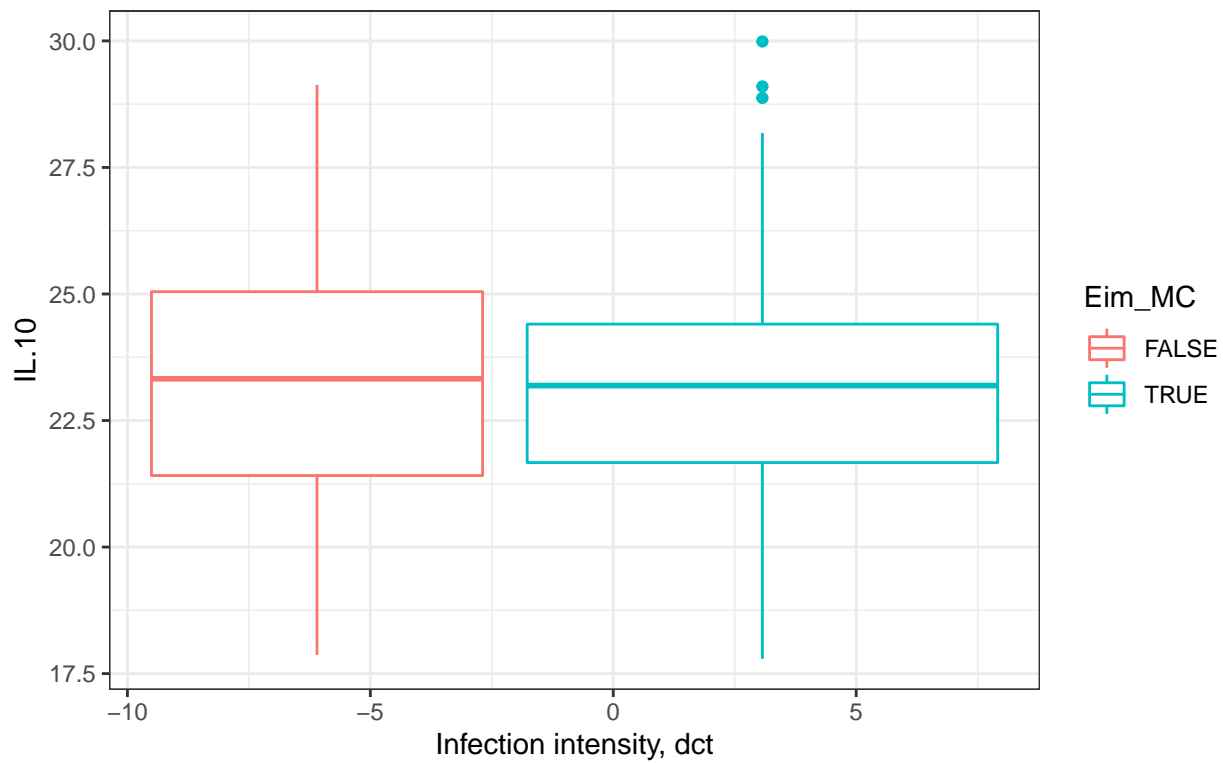## Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = IL.10, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

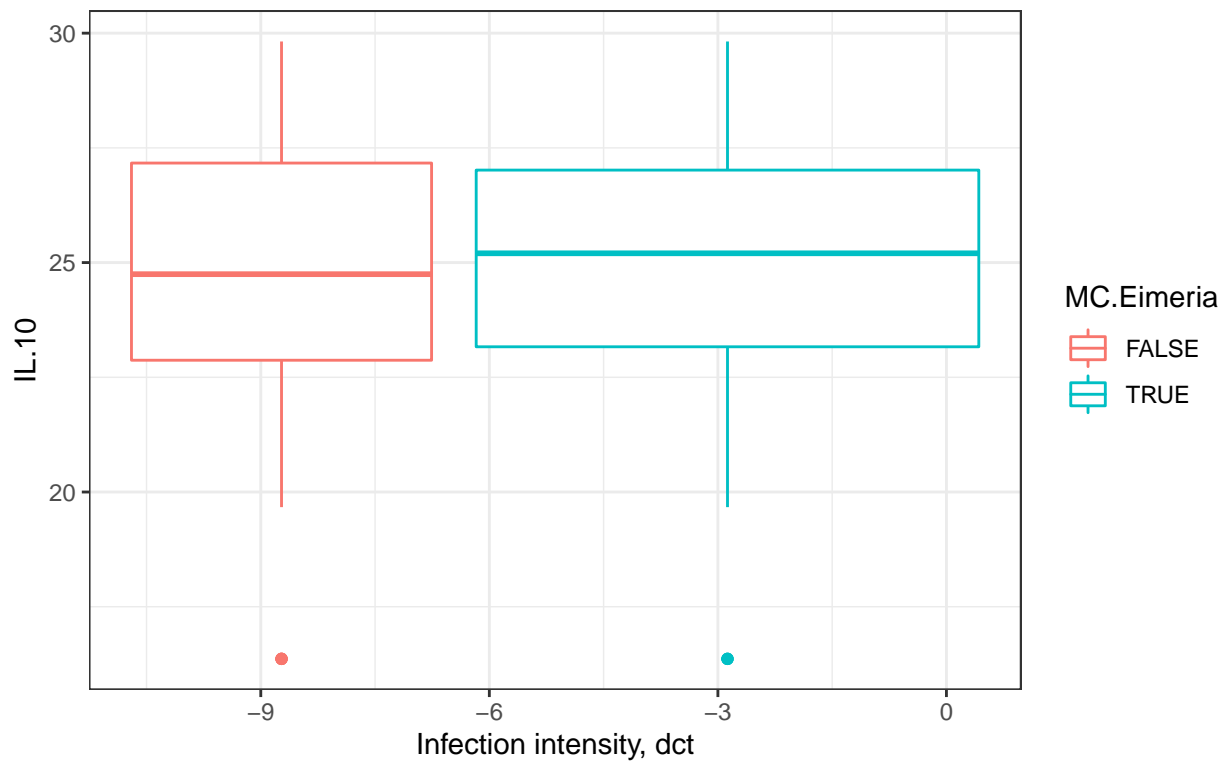## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL.10,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples
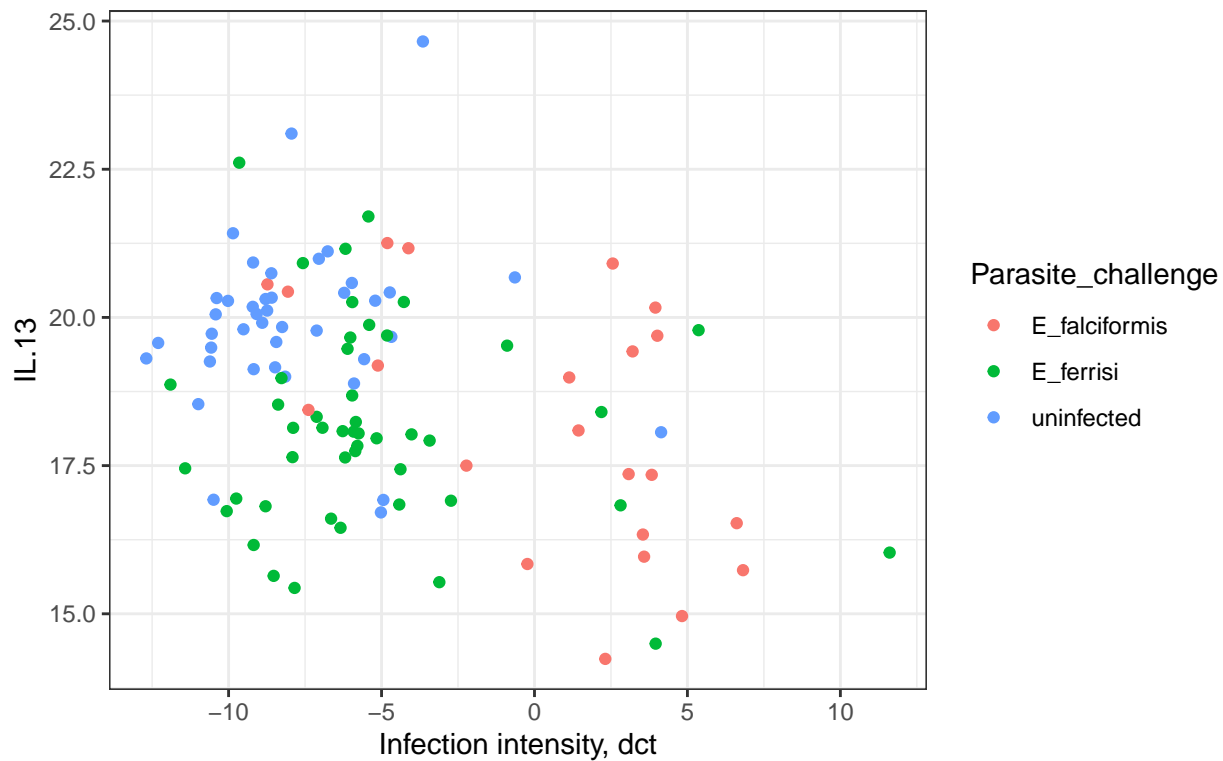

5. IL.13

```
ggplot(lab, aes(x = delta, y = IL.13, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

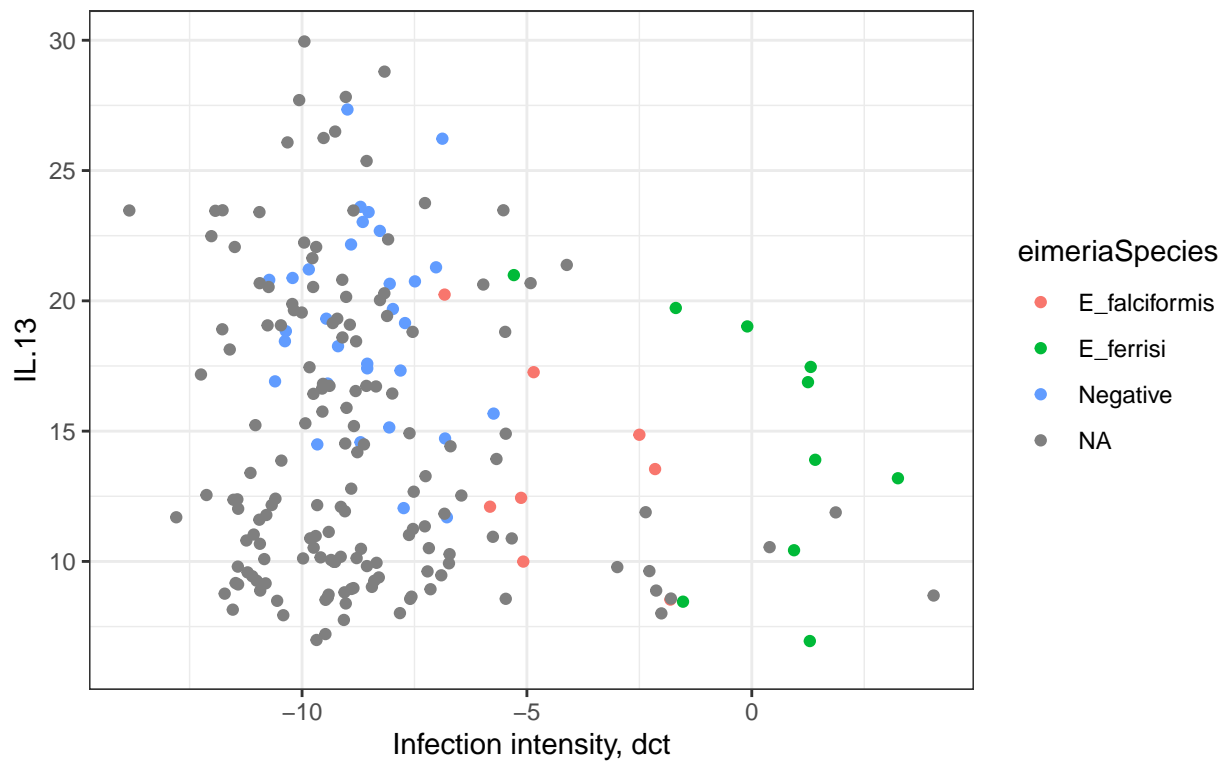# Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL.13, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```
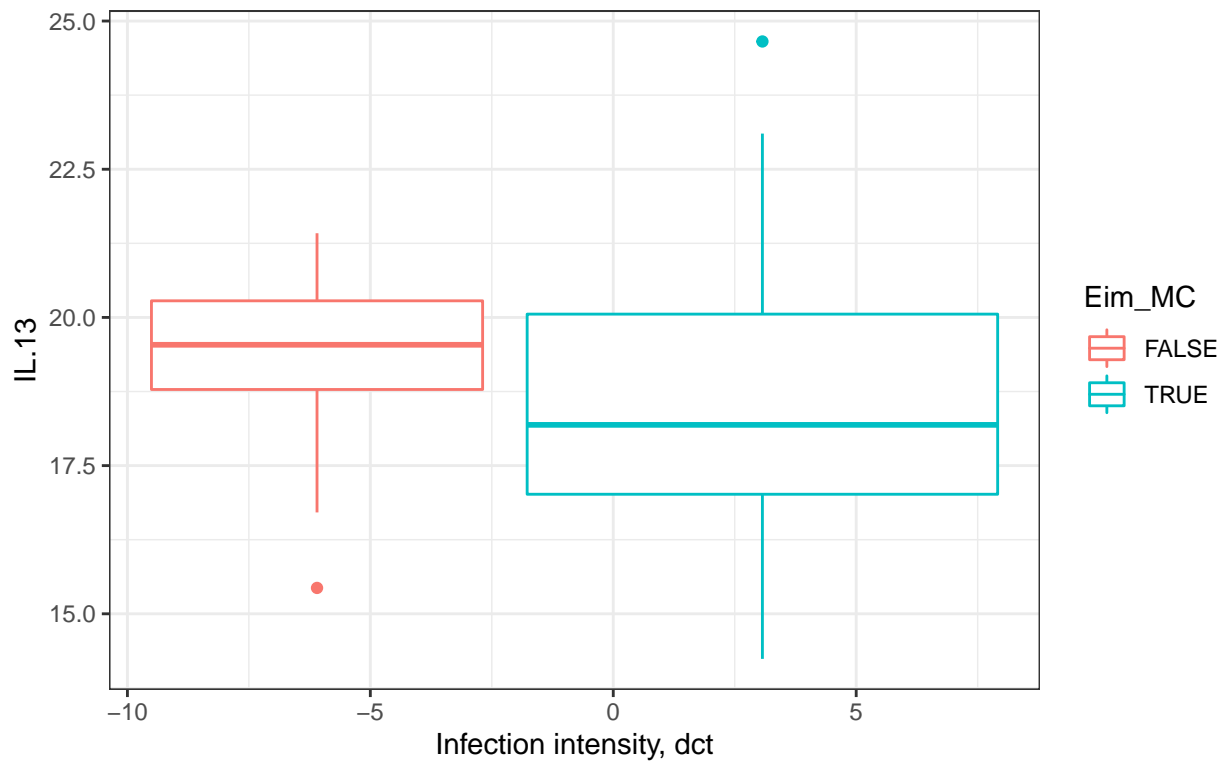
Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = IL.13, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

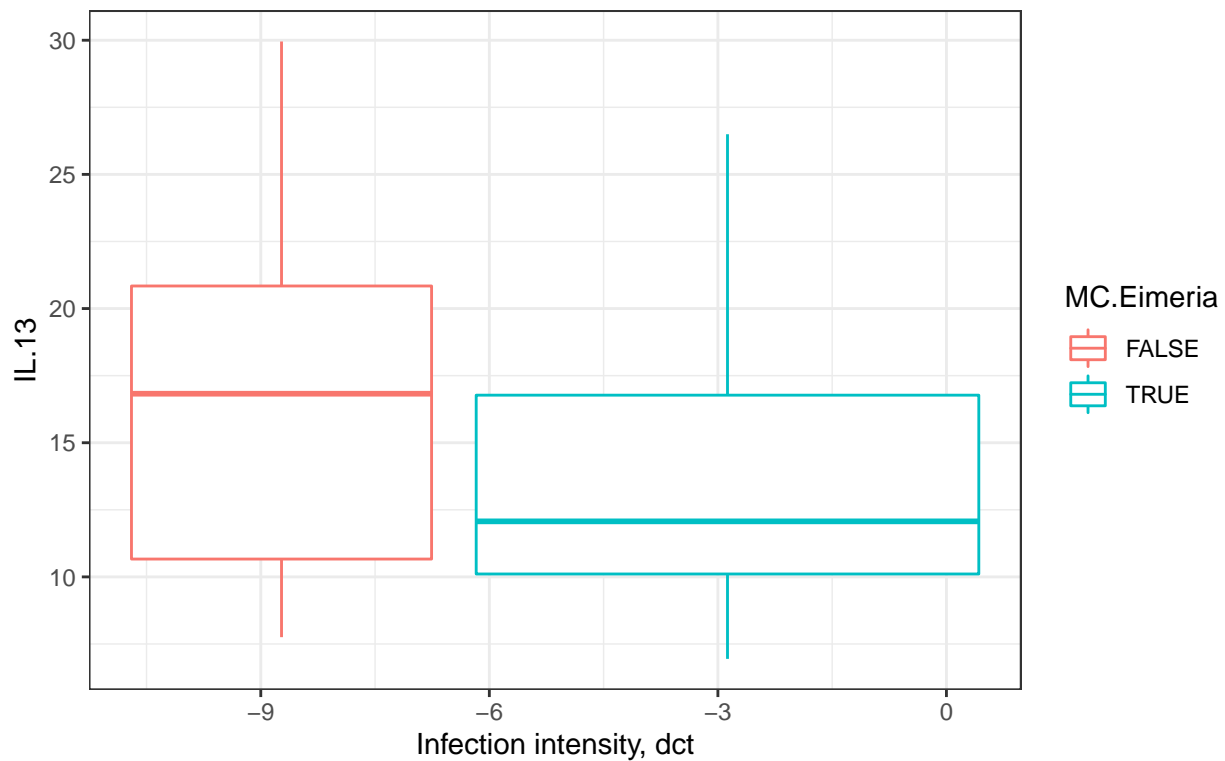## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections

```r
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL.13,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (stat_boxplot).
```

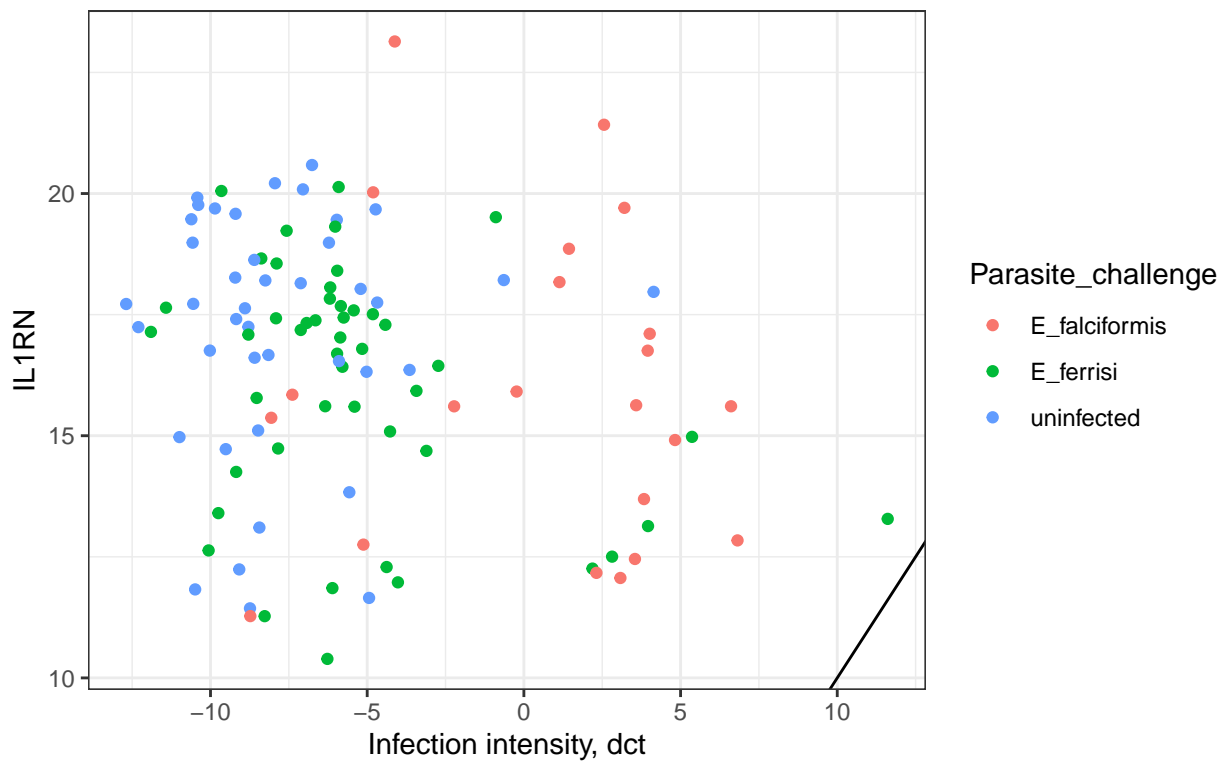Immune gene expression against infection intensity in field samples

6. IL1RN

```
ggplot(lab, aes(x = delta, y = IL1RN, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

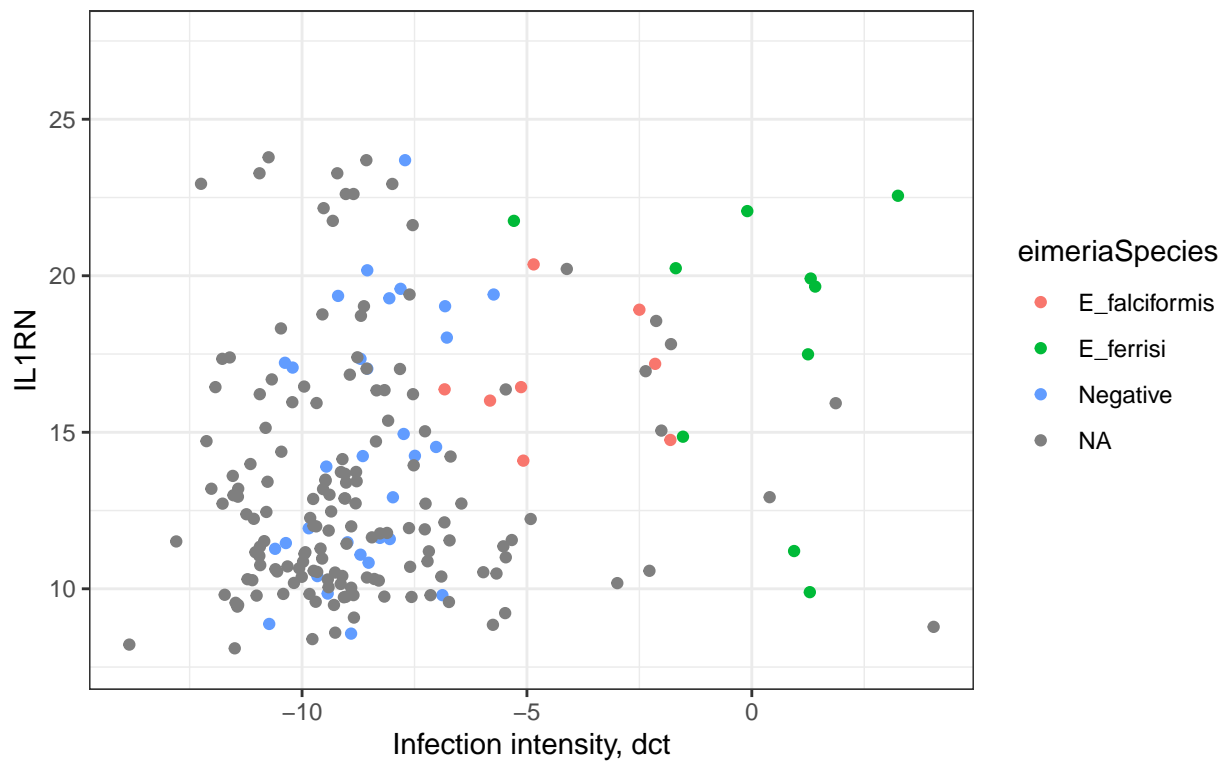## Warning: Removed 6 rows containing missing values (geom_point).

Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL1RN, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```
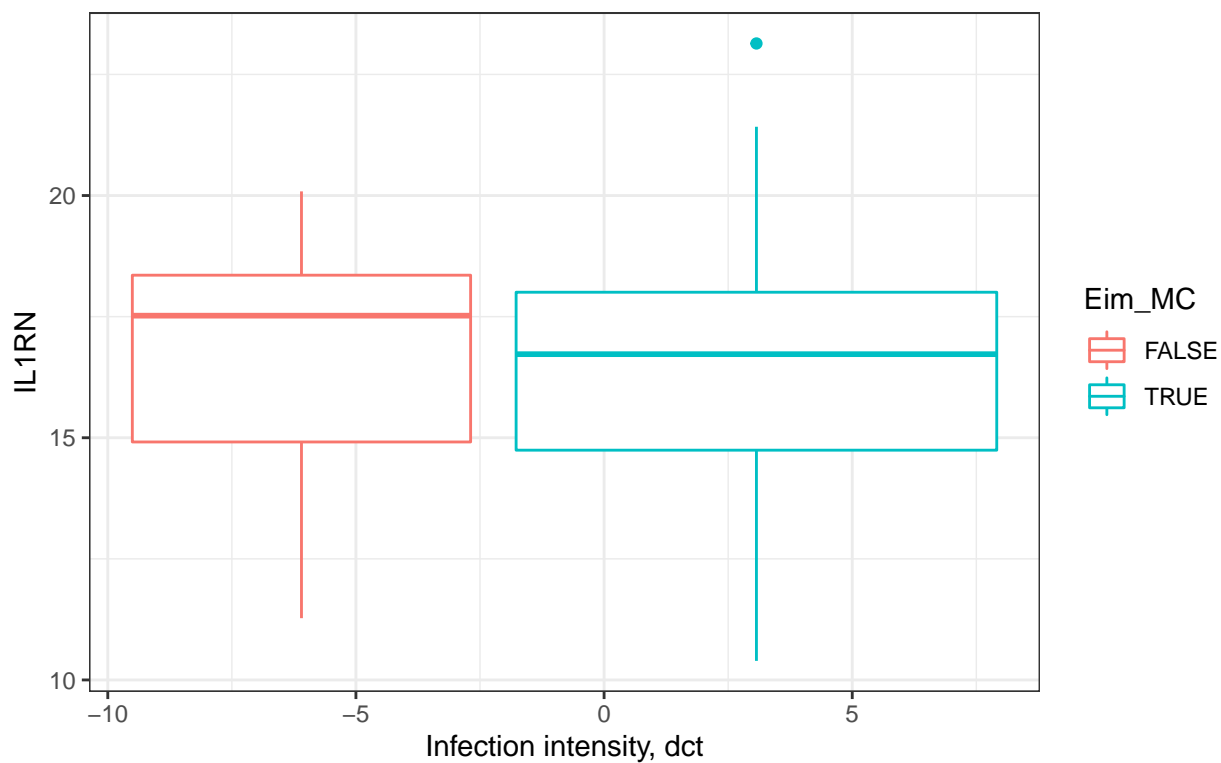
Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = IL1RN, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IL1RN,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

7. CASP1

```
ggplot(lab, aes(x = delta, y = CASP1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

Warning: Removed 6 rows containing missing values (geom_point).
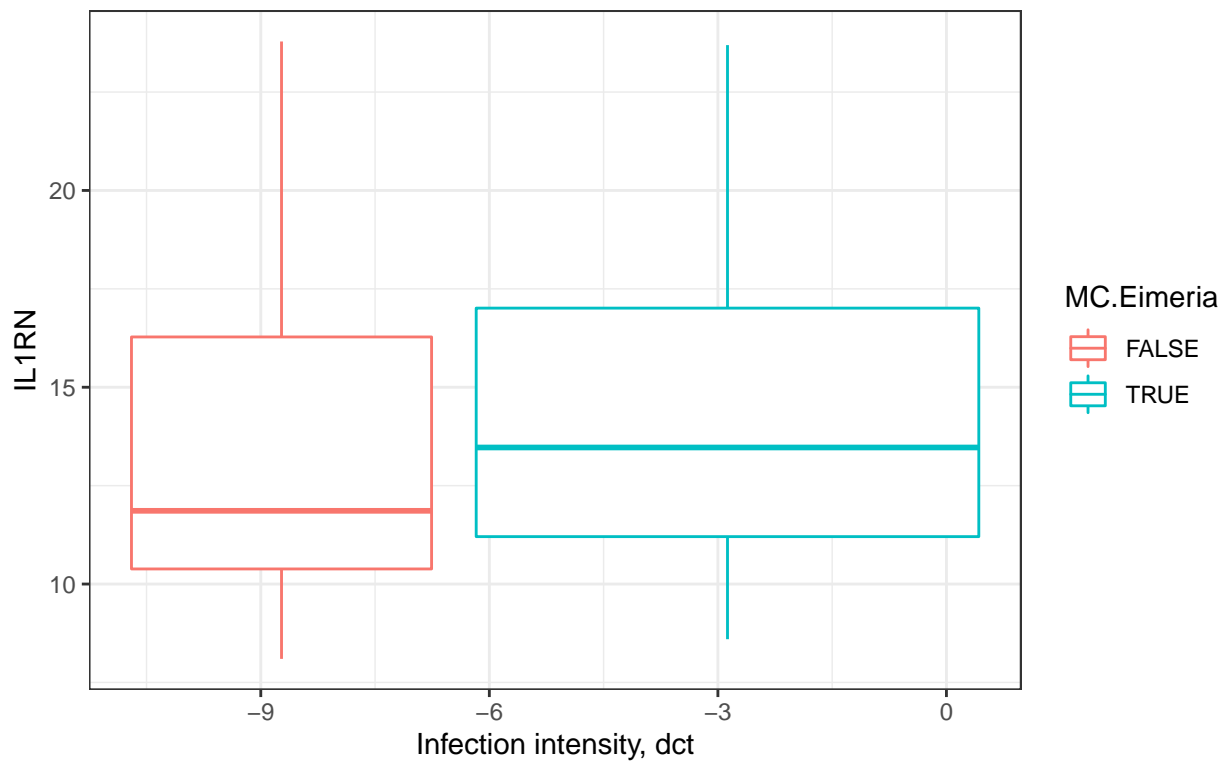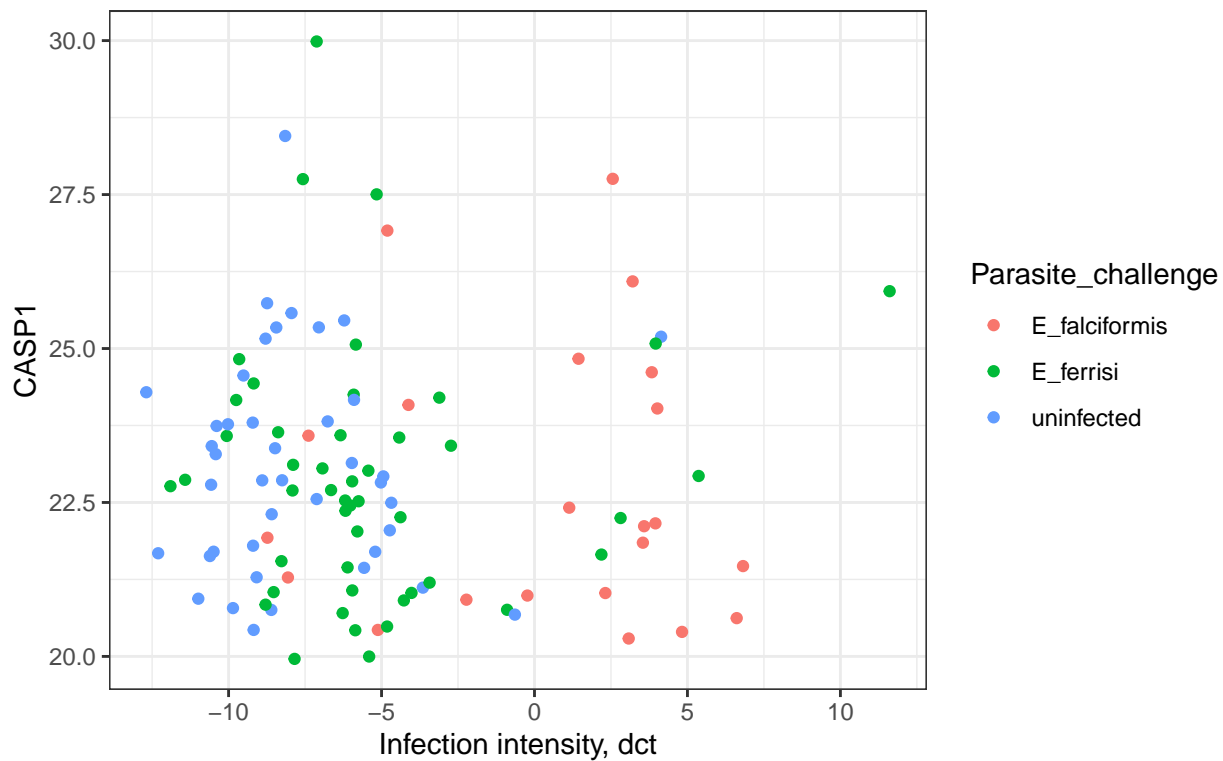
Immune gene expression against infection intensity in lab infections
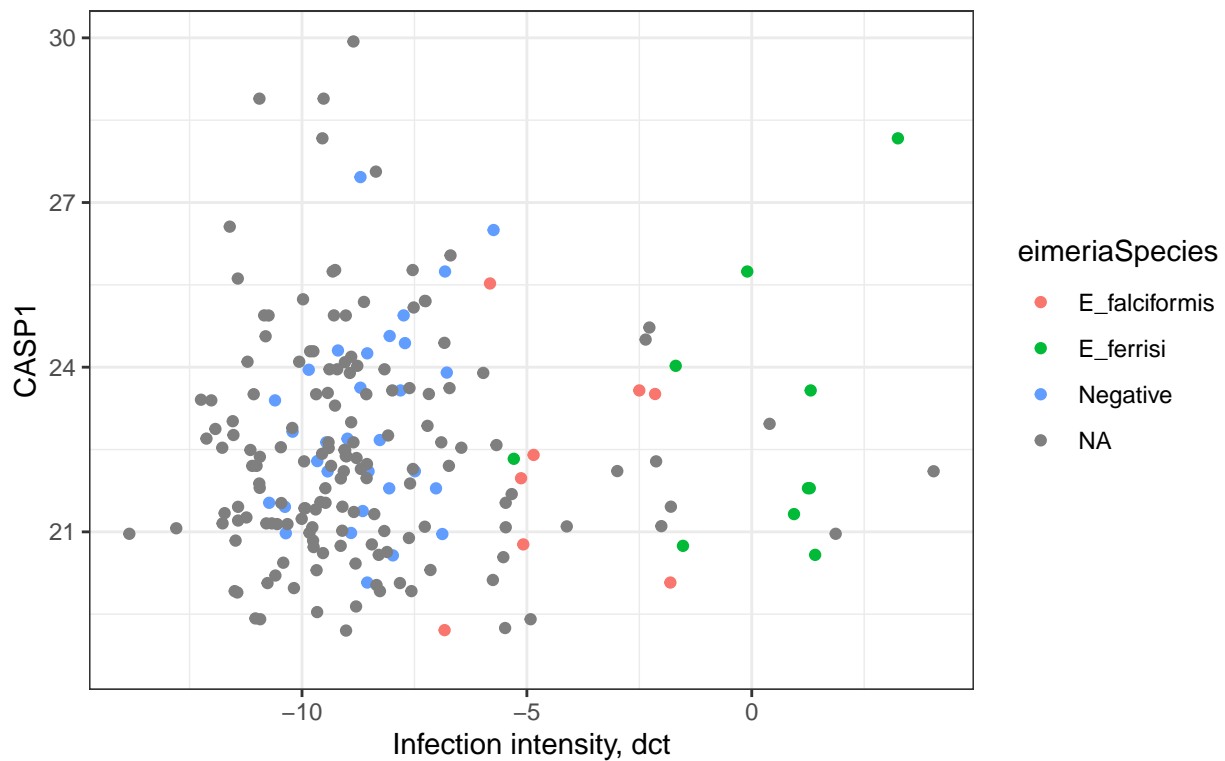


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = CASP1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (geom_point).
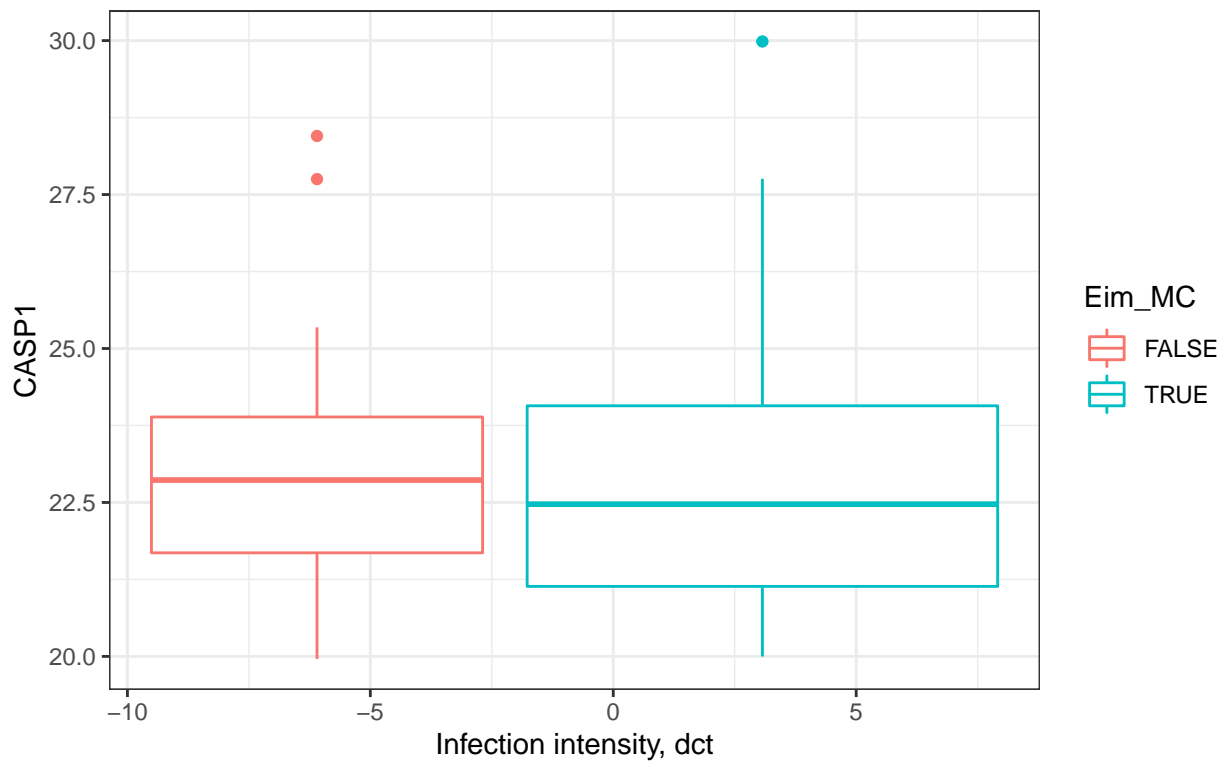
Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = CASP1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = CASP1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

8. CXCL9

```
ggplot(lab, aes(x = delta, y = CXCL9, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

Warning: Removed 6 rows containing missing values (geom_point).
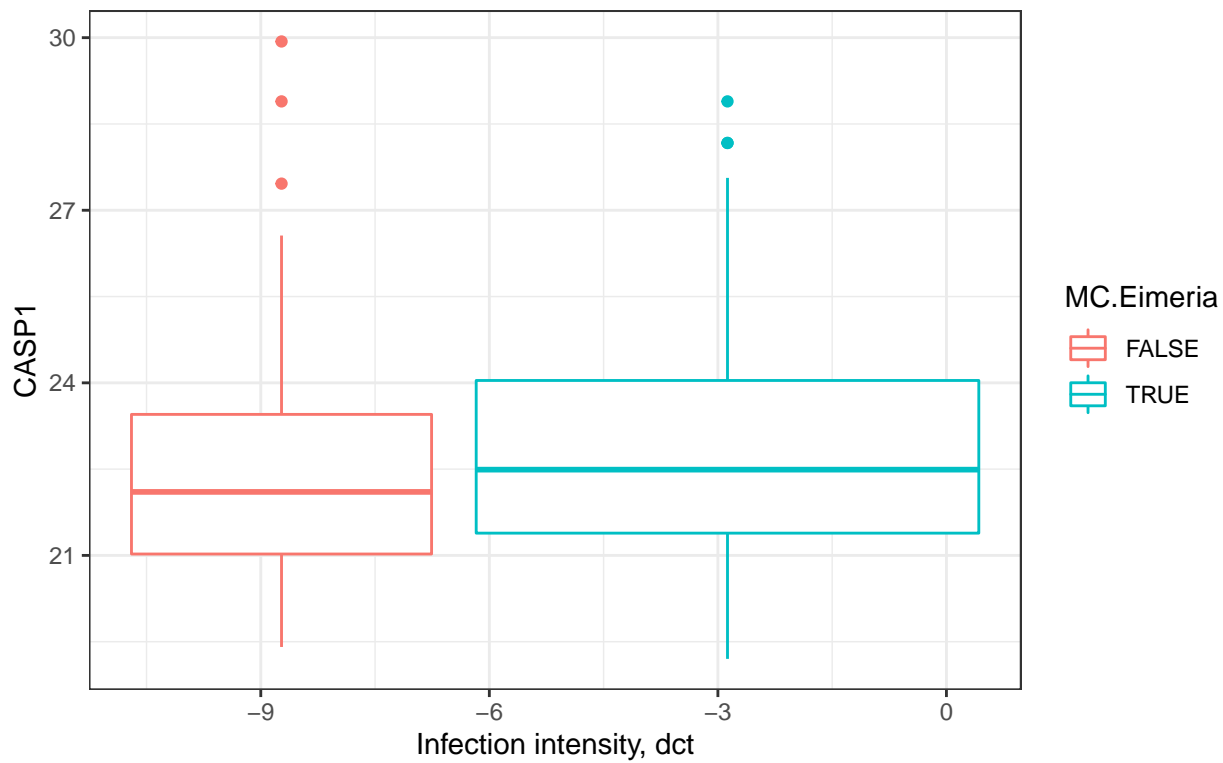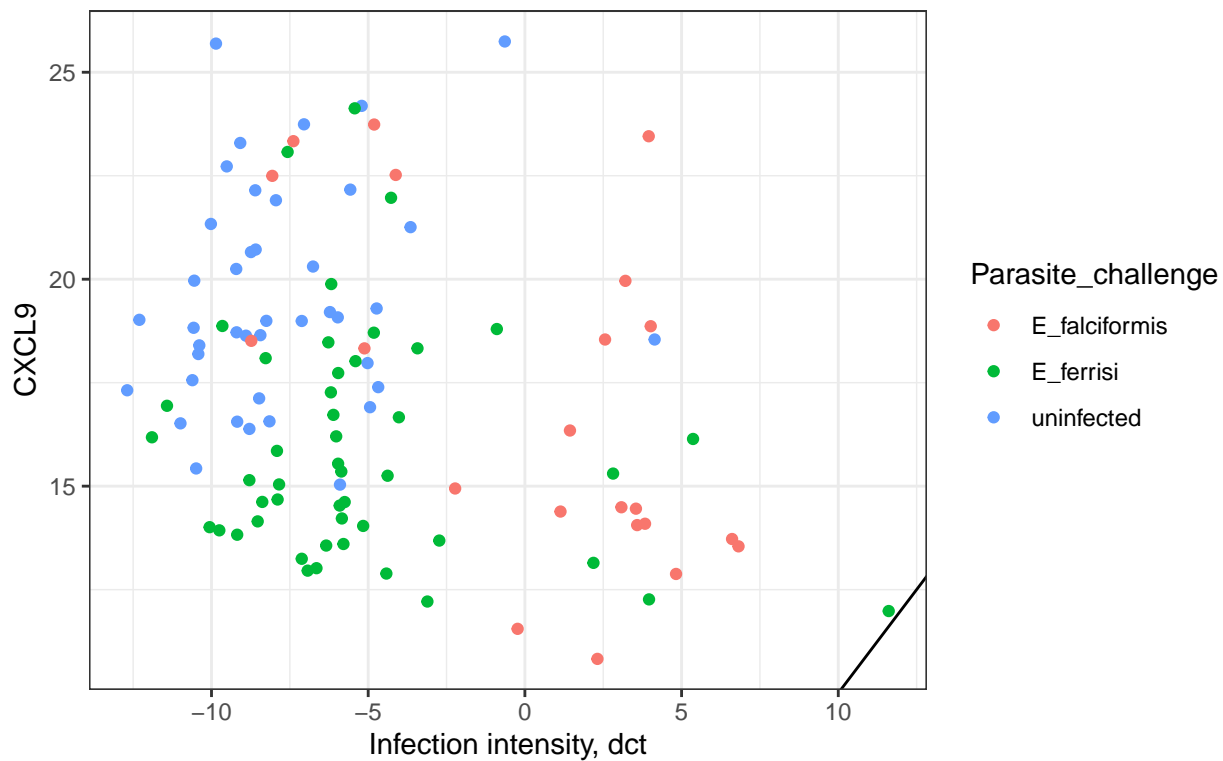
Immune gene expression against infection intensity
in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = CXCL9, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples

```
ggplot(lab, aes(x = delta, y = CXCL9, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

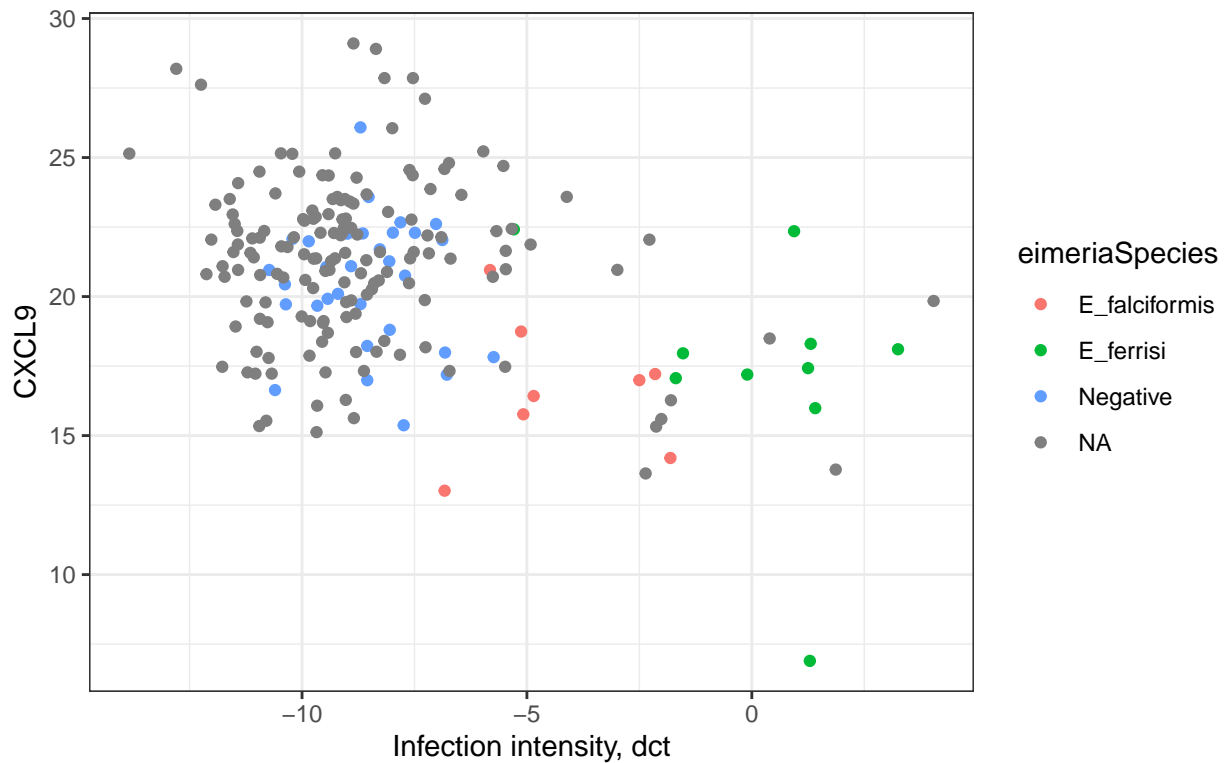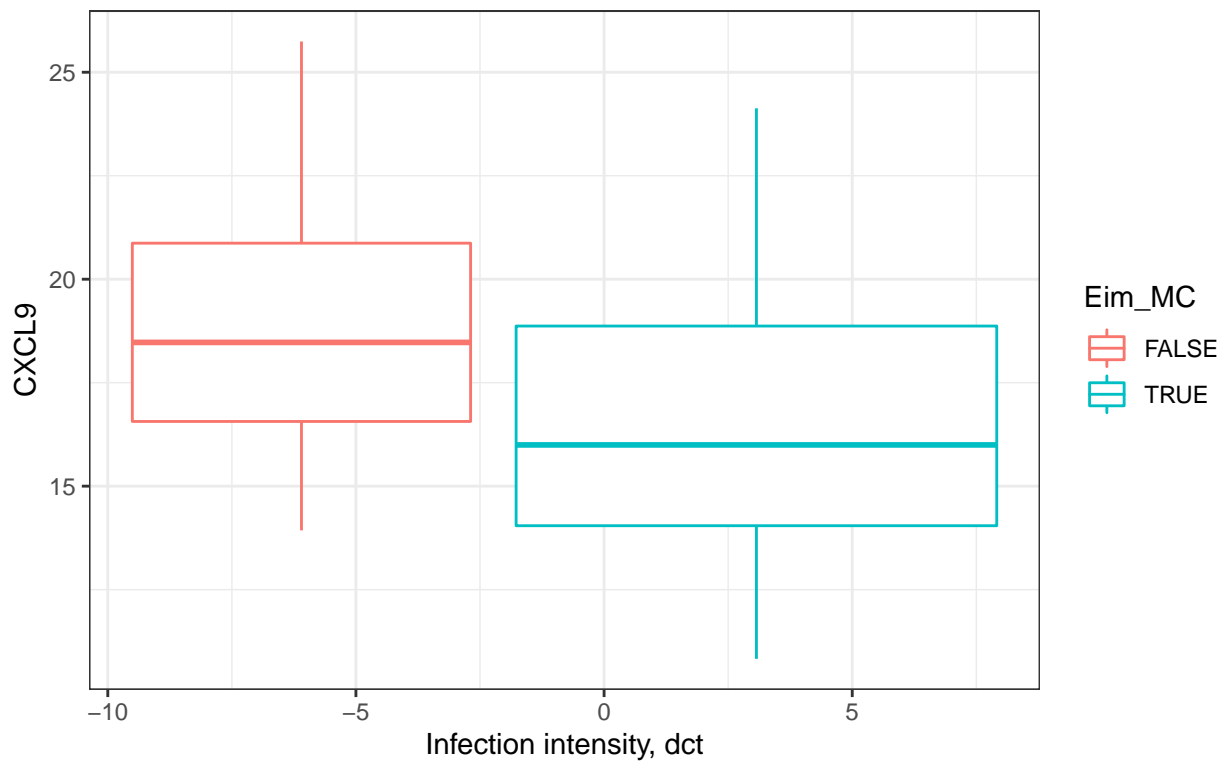## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = CXCL9,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

9. IDO1

```
ggplot(lab, aes(x = delta, y = IDO1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
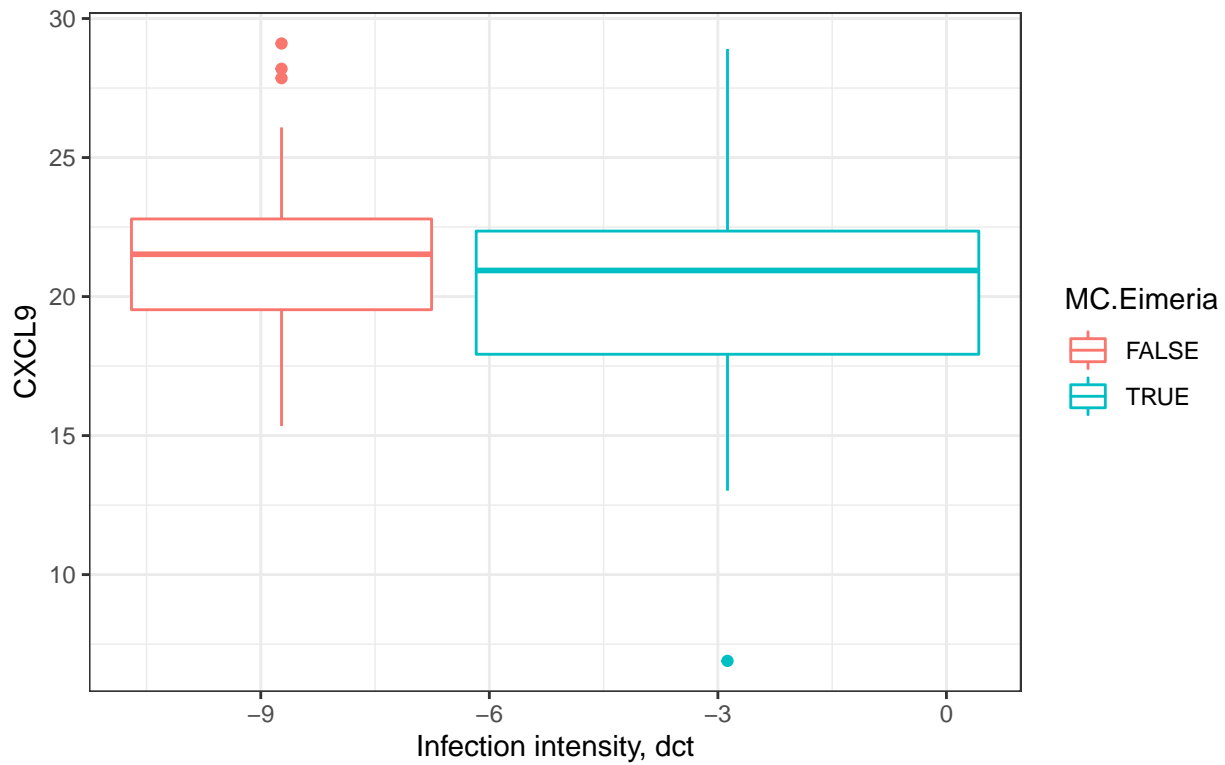
Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IDO1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```

Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = IDO1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

Warning: Removed 6 rows containing missing values (stat_boxplot).

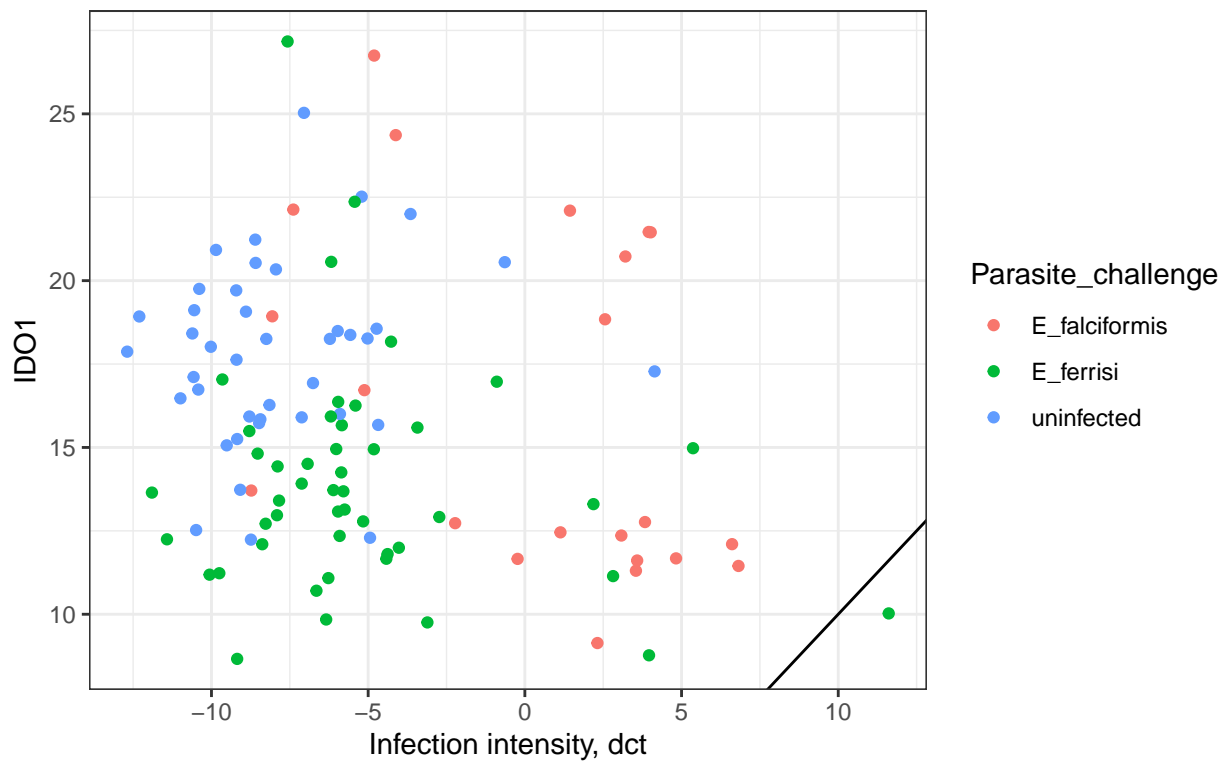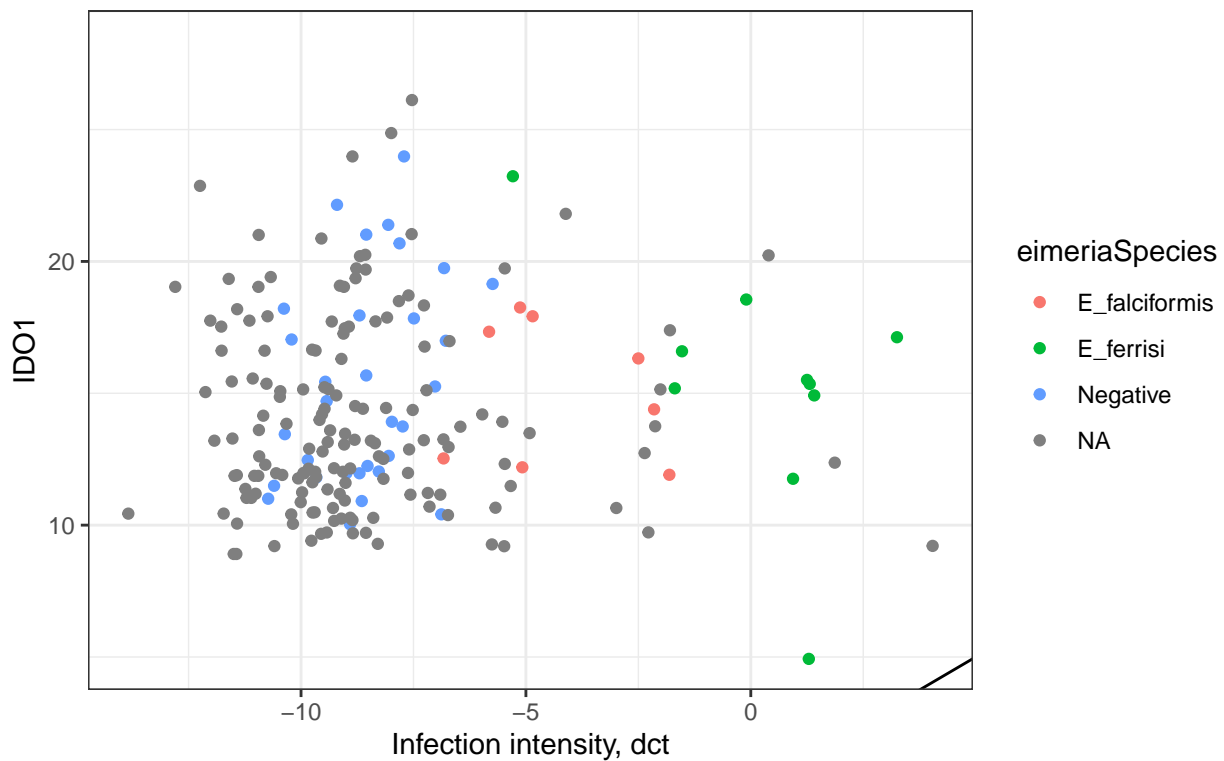## Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IDO1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

10. IRGM1

```
ggplot(lab, aes(x = delta, y = IRGM1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
      title ="Immune gene expression against infection intensity
      in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IRGM1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

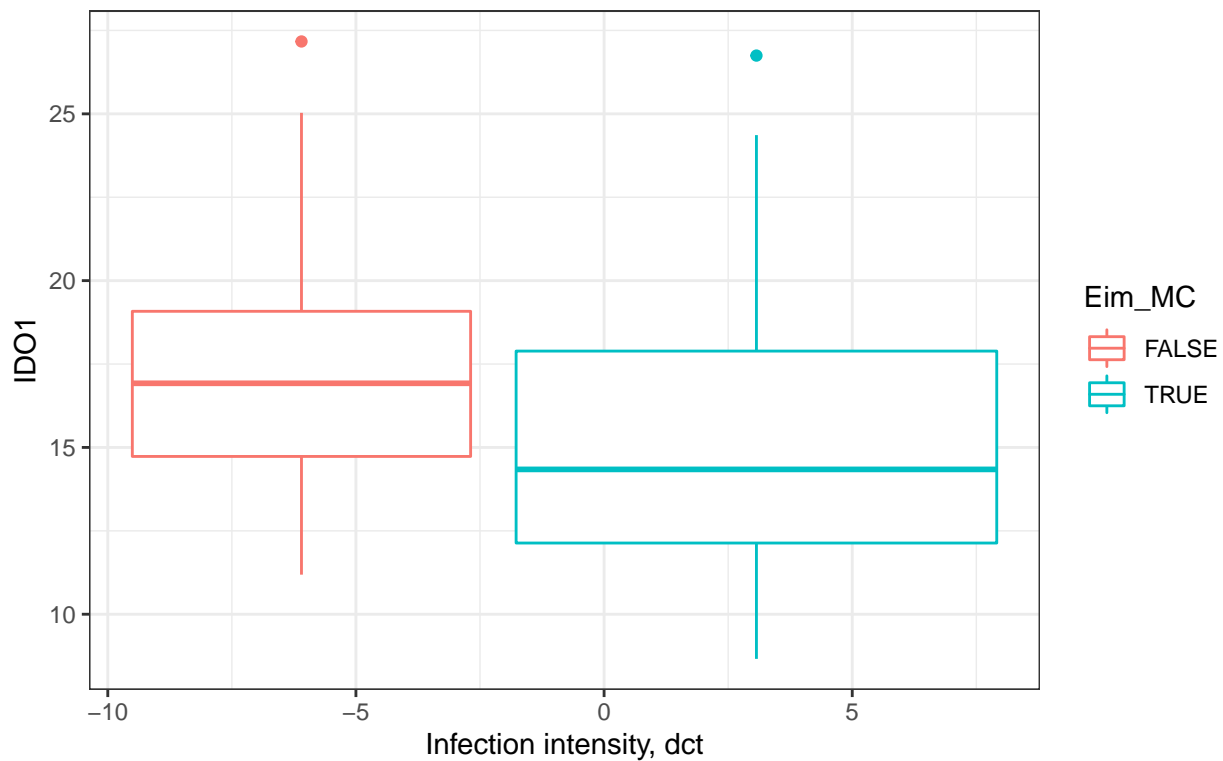## Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples

```
ggplot(lab, aes(x = delta, y = IRGM1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

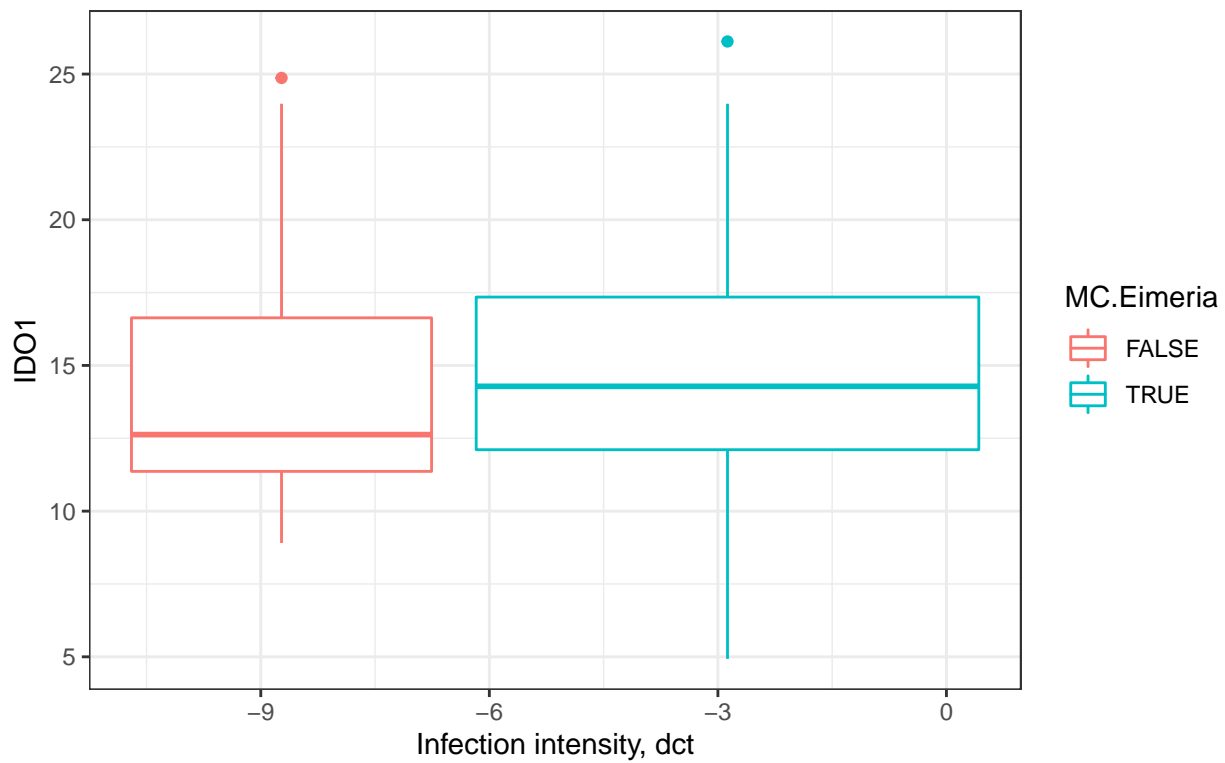## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections

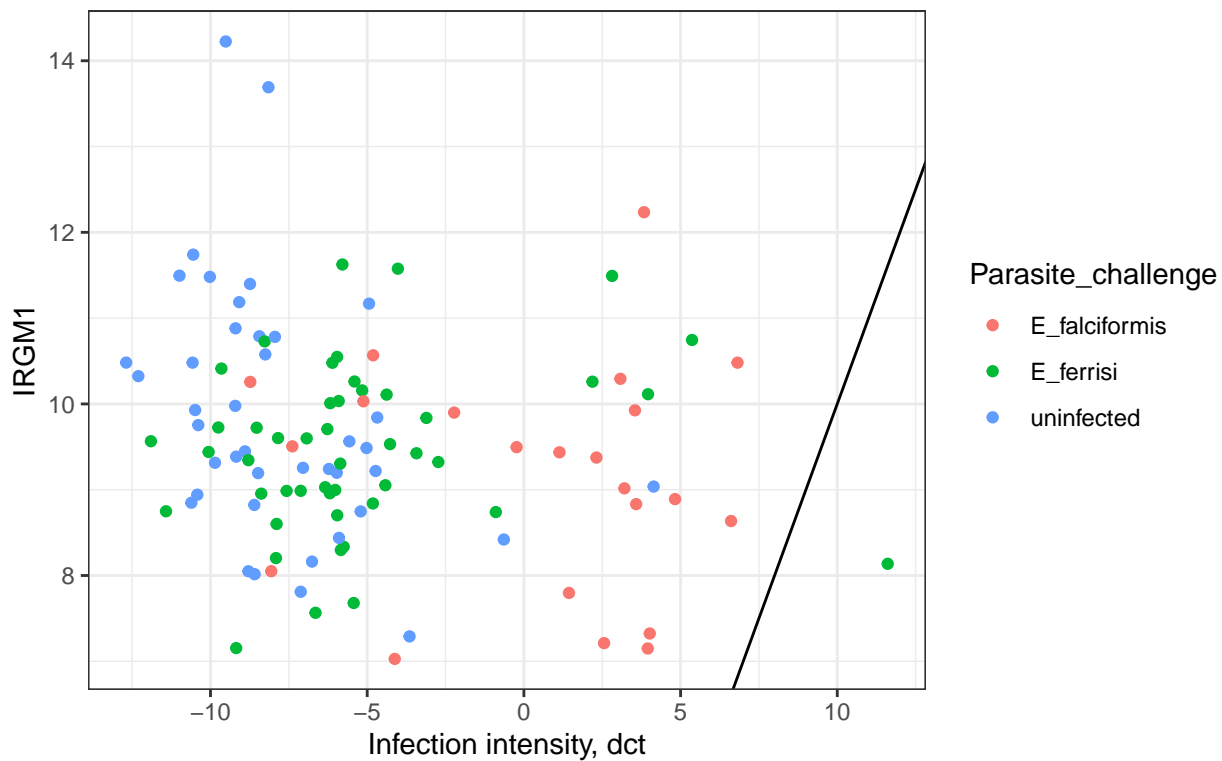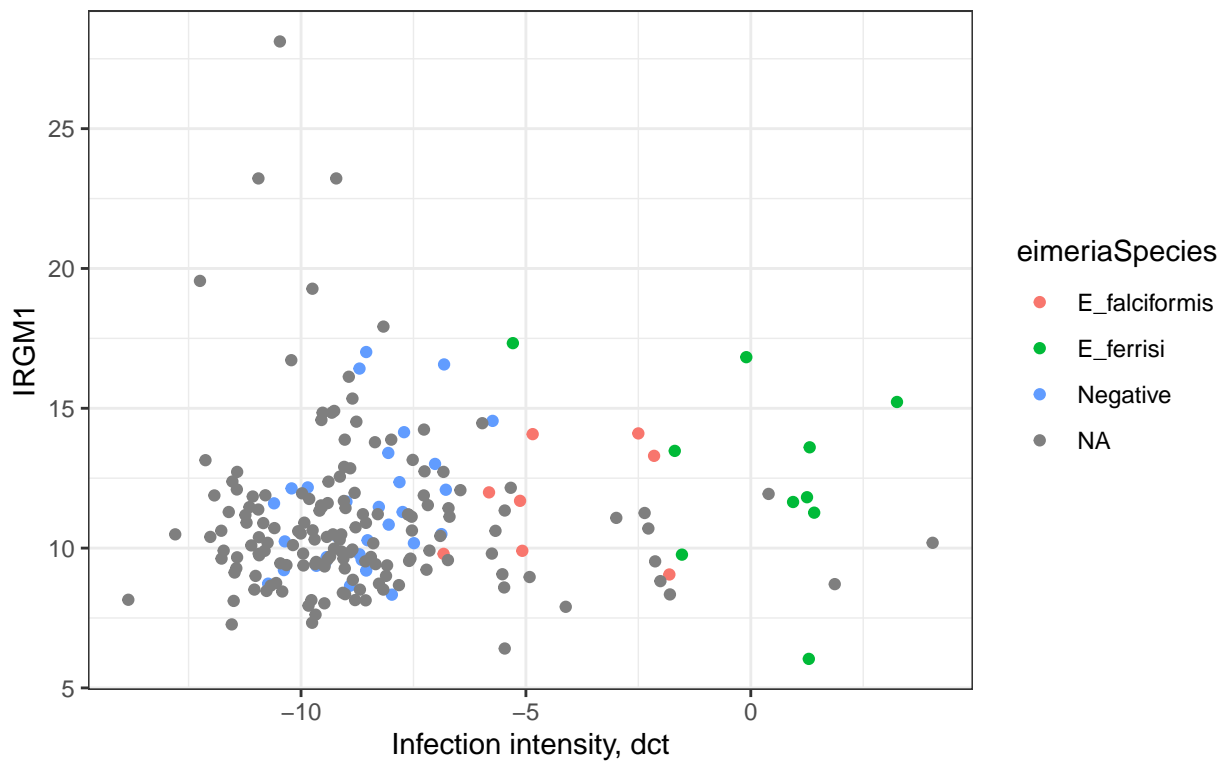

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = IRGM1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

11. MPO

```
ggplot(lab, aes(x = delta, y = MPO, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
      title ="Immune gene expression against infection intensity
      in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

Immune gene expression against infection intensity
in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MPO, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

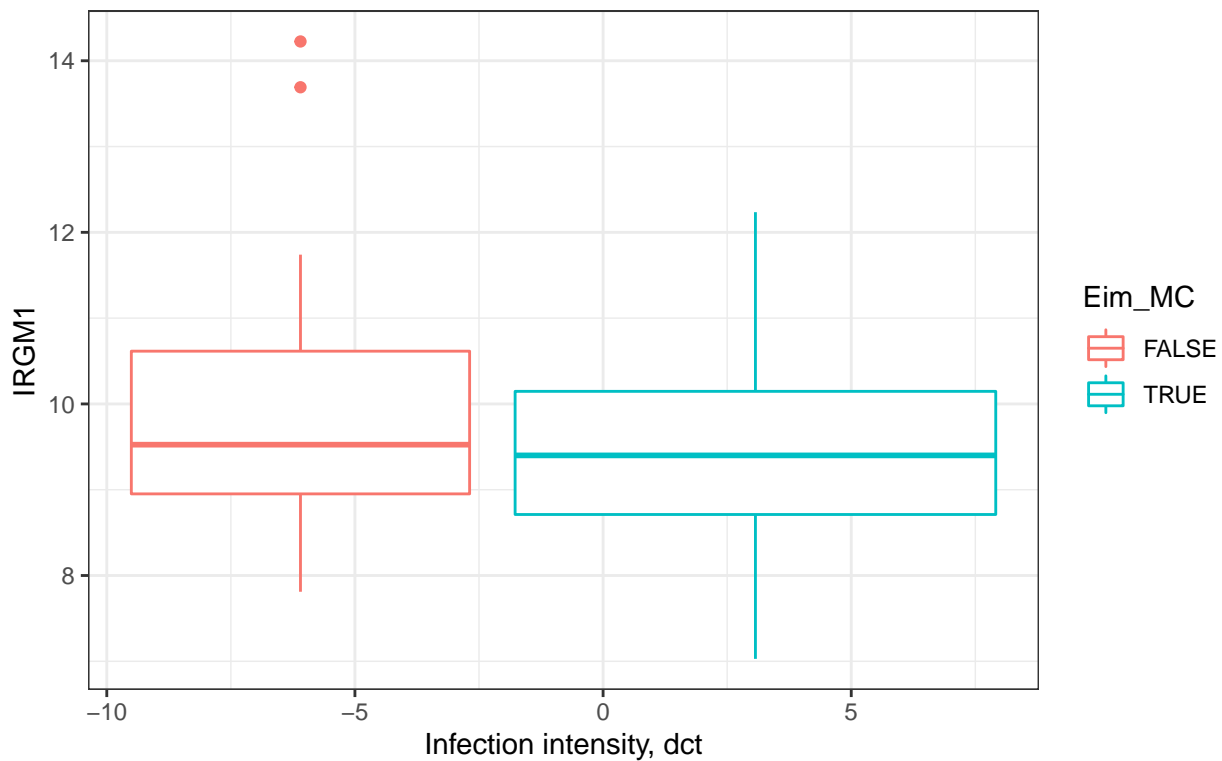## Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = MPO, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

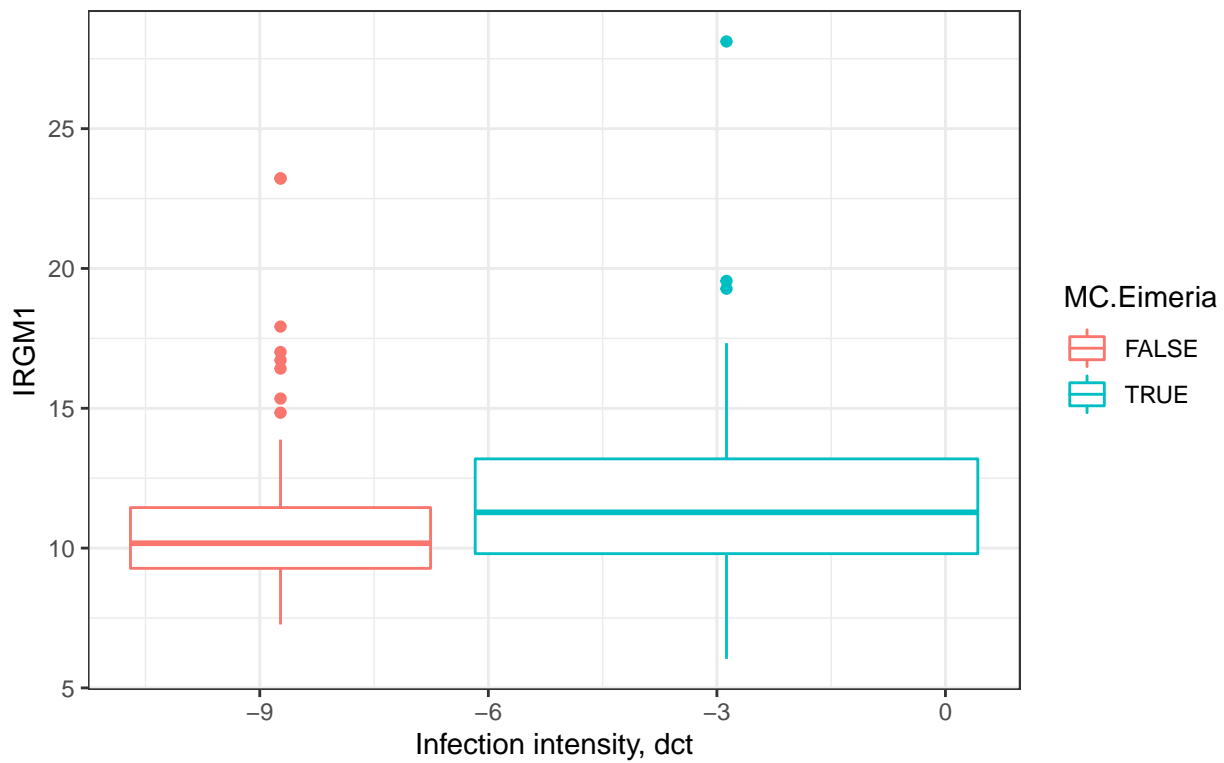## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MPO,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

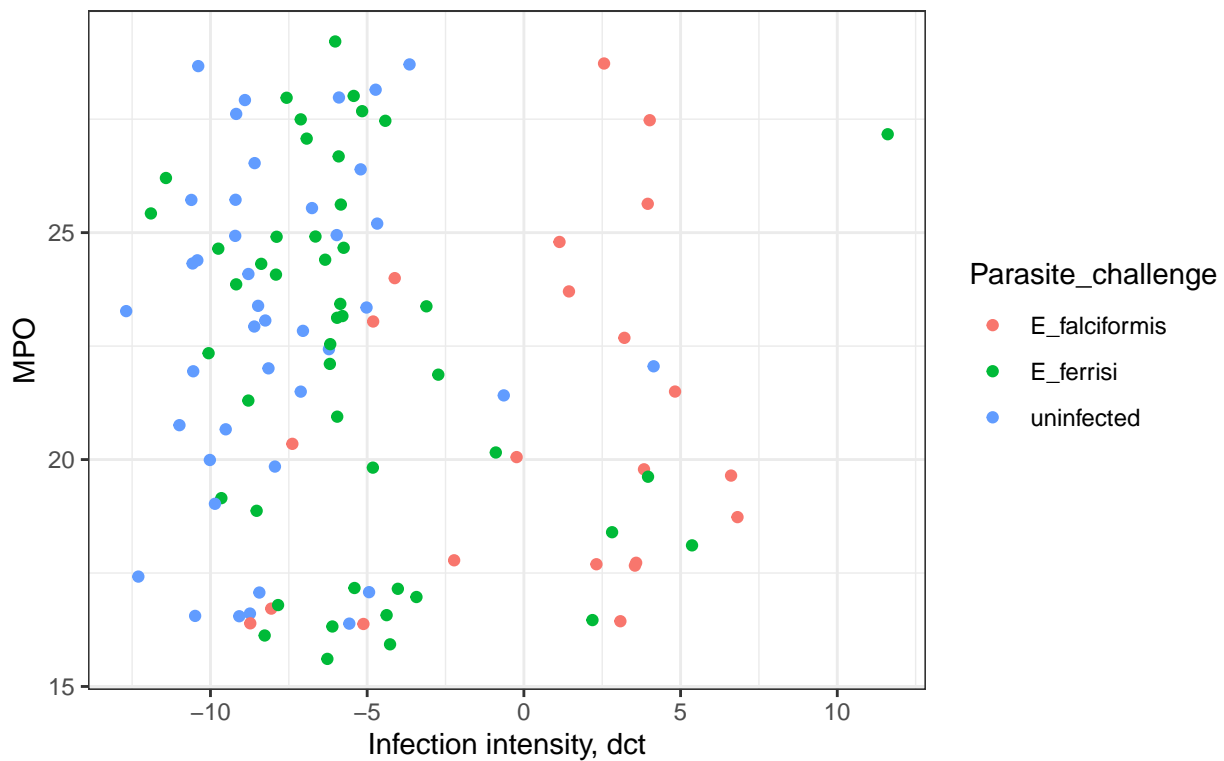Immune gene expression against infection intensity in field samples

12. MUC2

```
ggplot(lab, aes(x = delta, y = MUC2, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

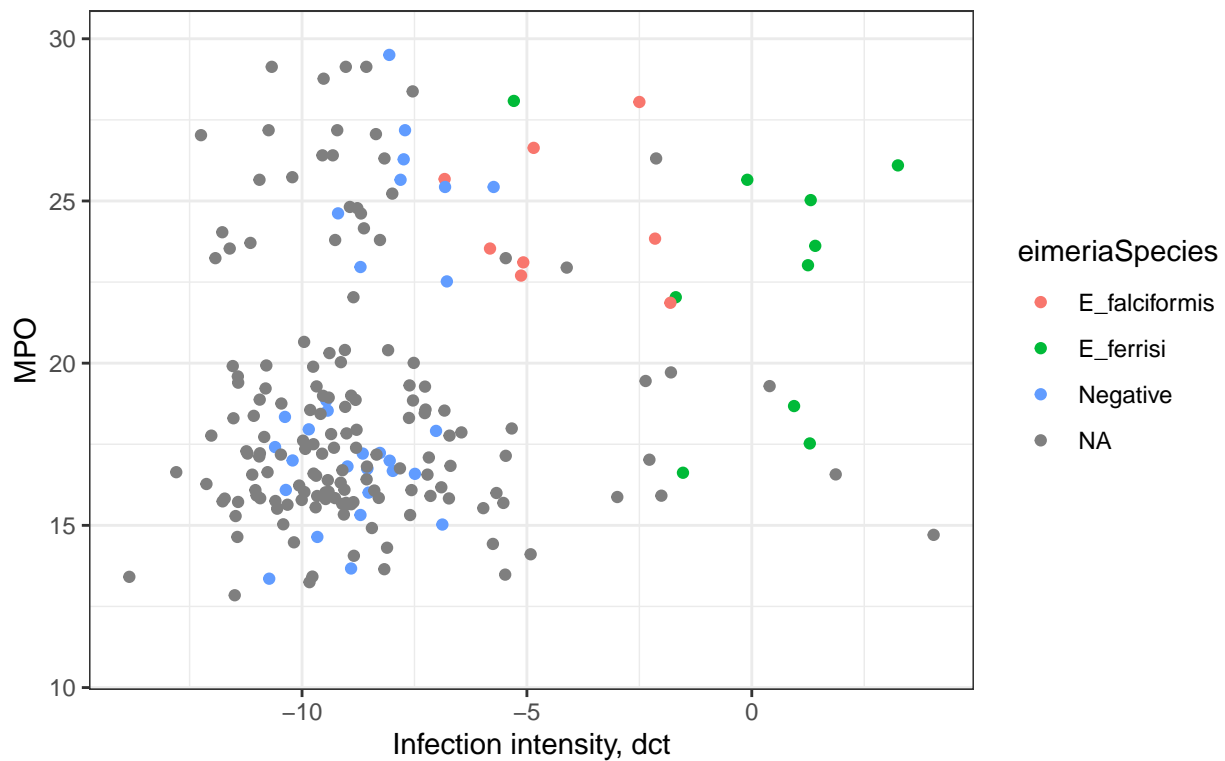## Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MUC2, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

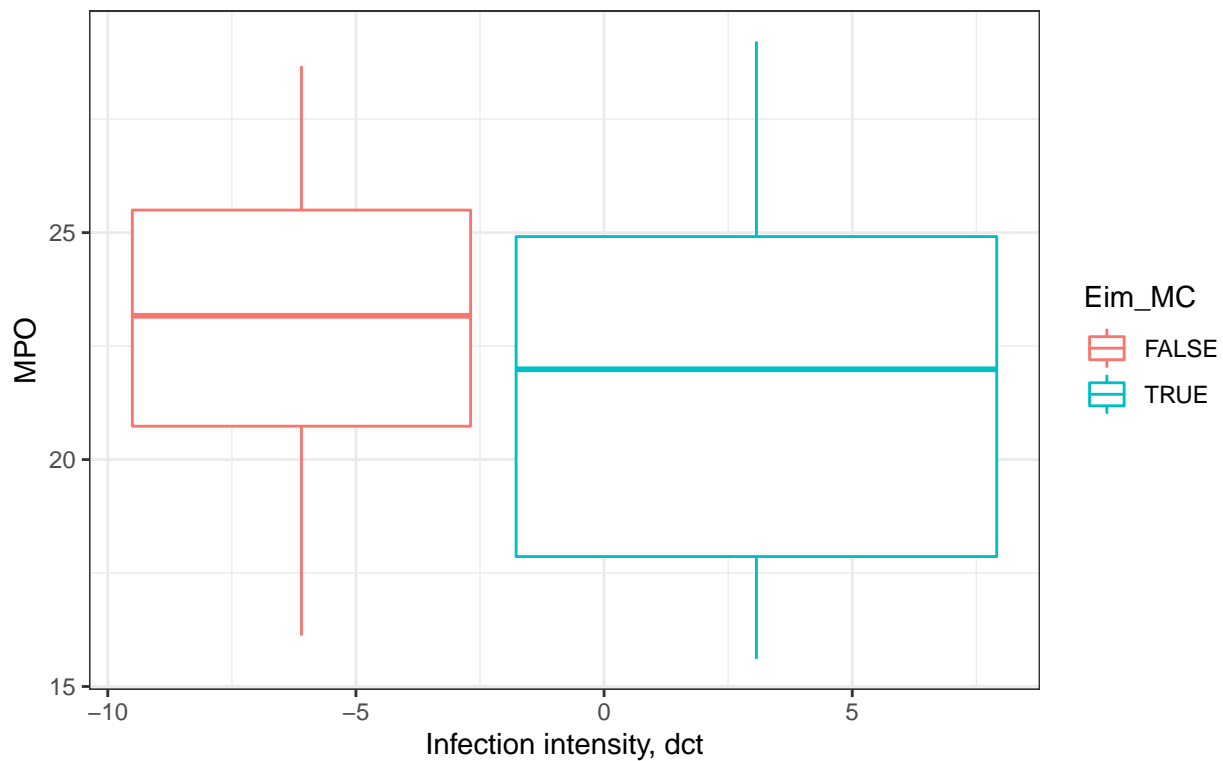## Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = MUC2, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

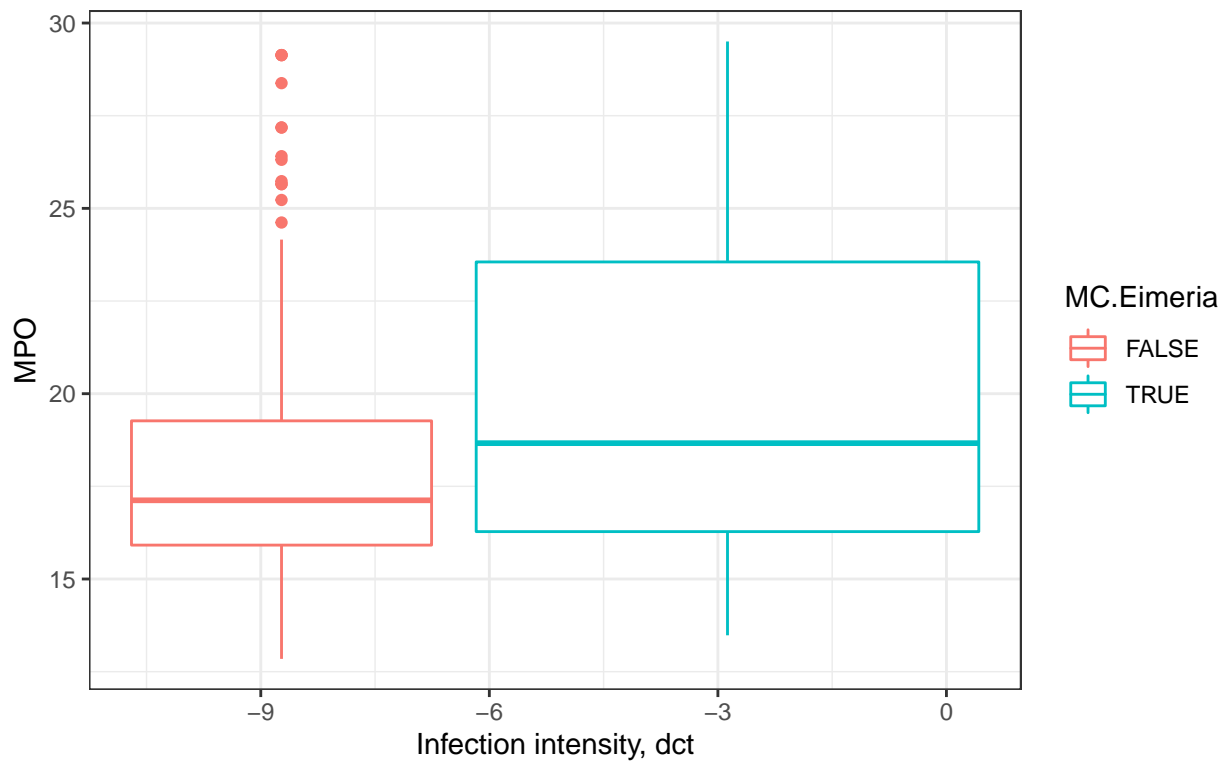## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MUC2,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

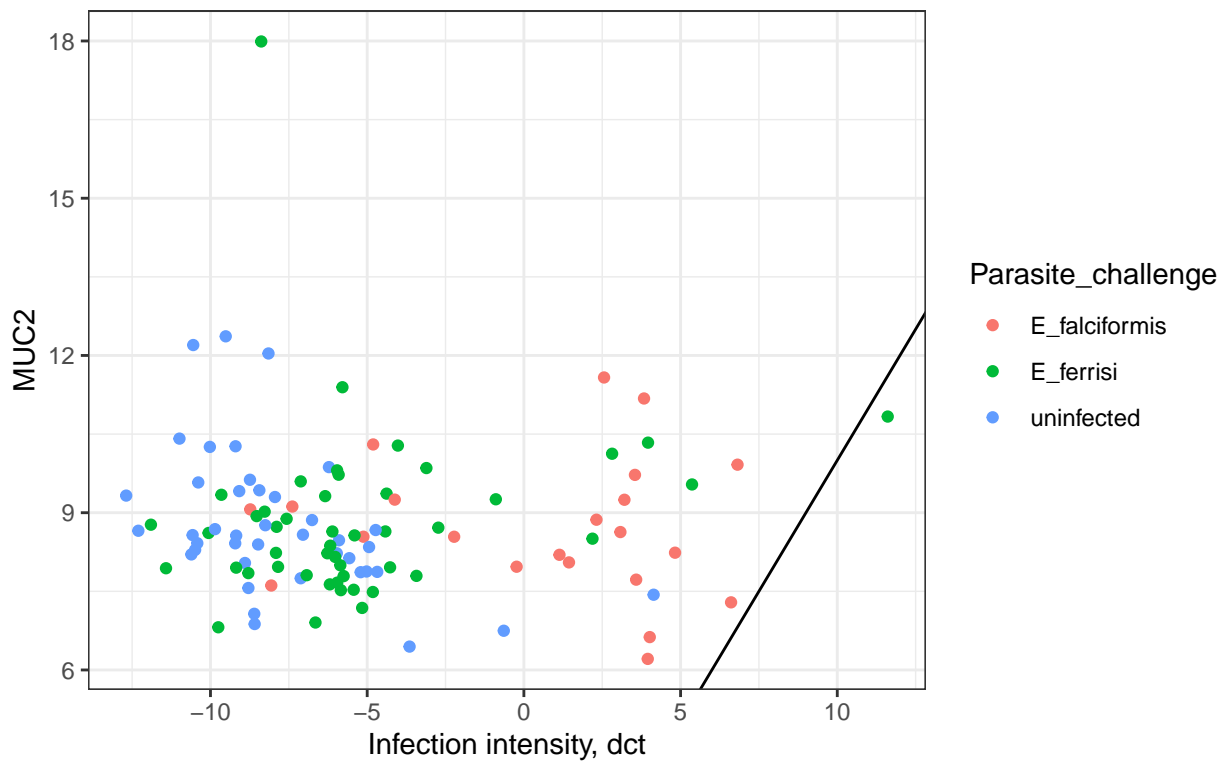Immune gene expression against infection intensity in field samples

13. MUC5AC

```
ggplot(lab, aes(x = delta, y = MUC5AC, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
      title ="Immune gene expression against infection intensity
      in lab infections") +
  theme_bw()
```

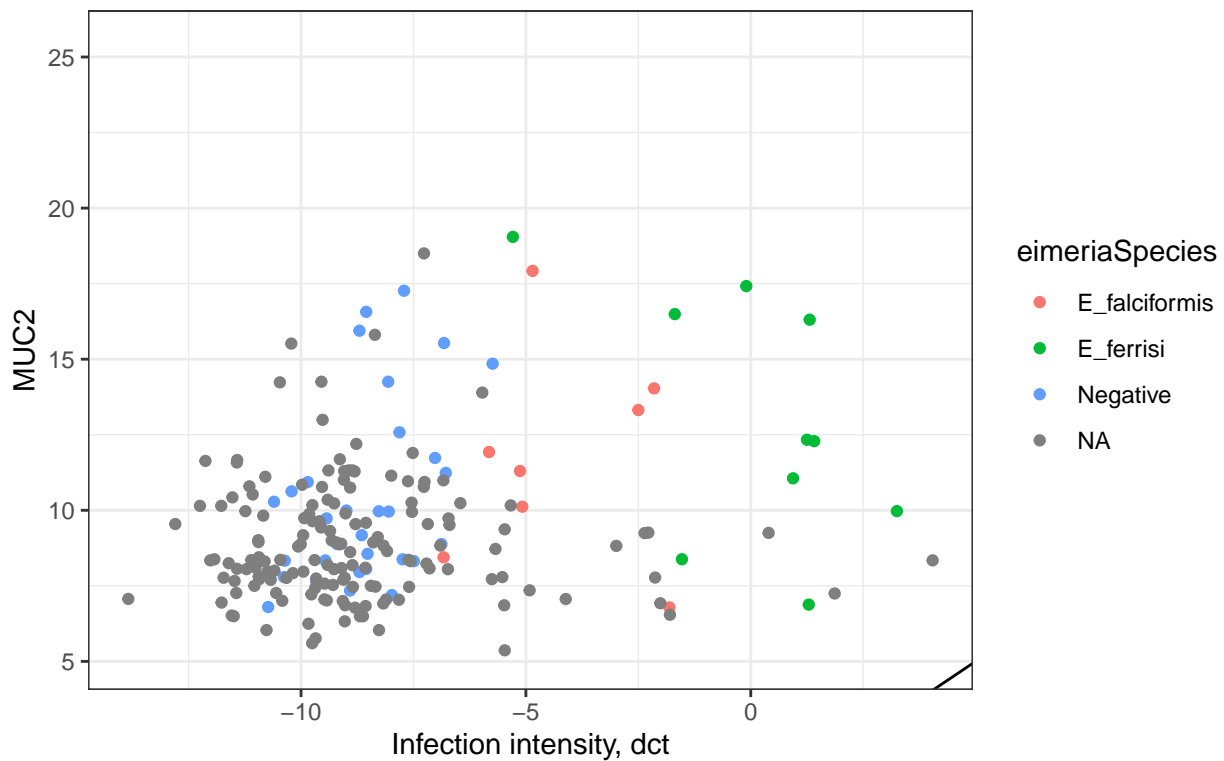## Warning: Removed 6 rows containing missing values (geom_point).

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MUC5AC, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```
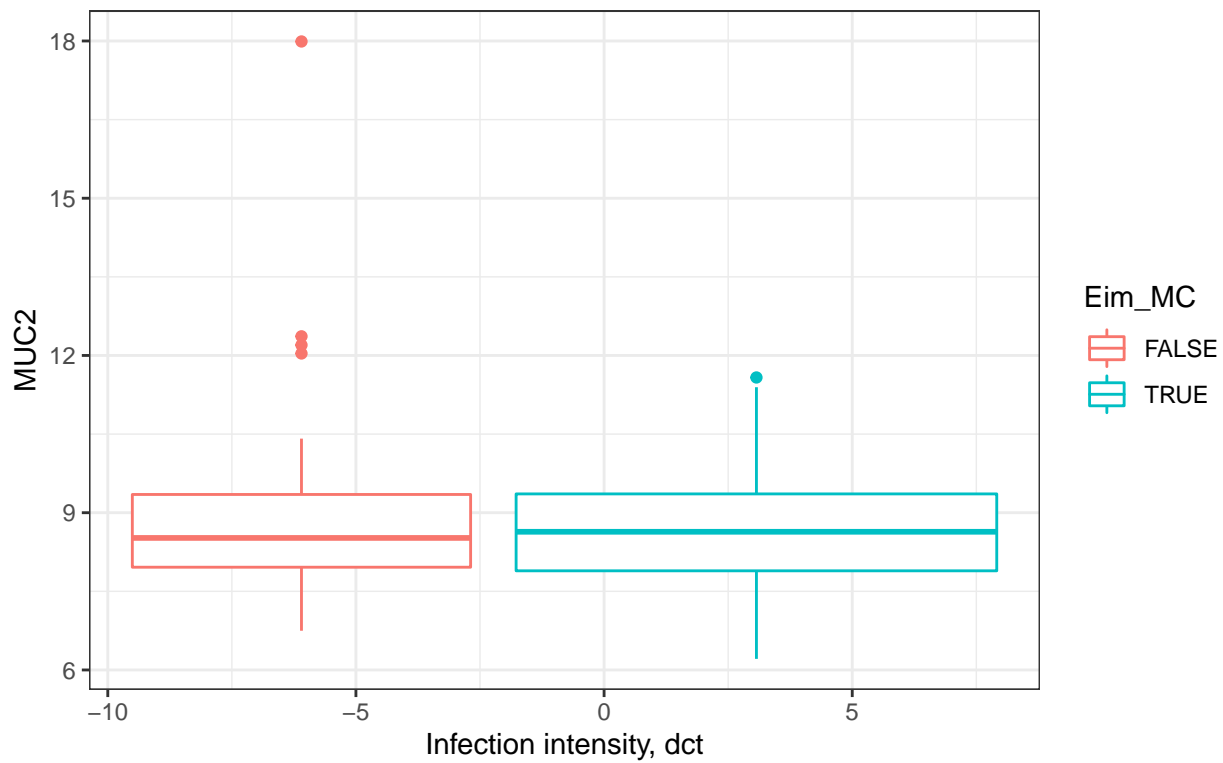
Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = MUC5AC, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

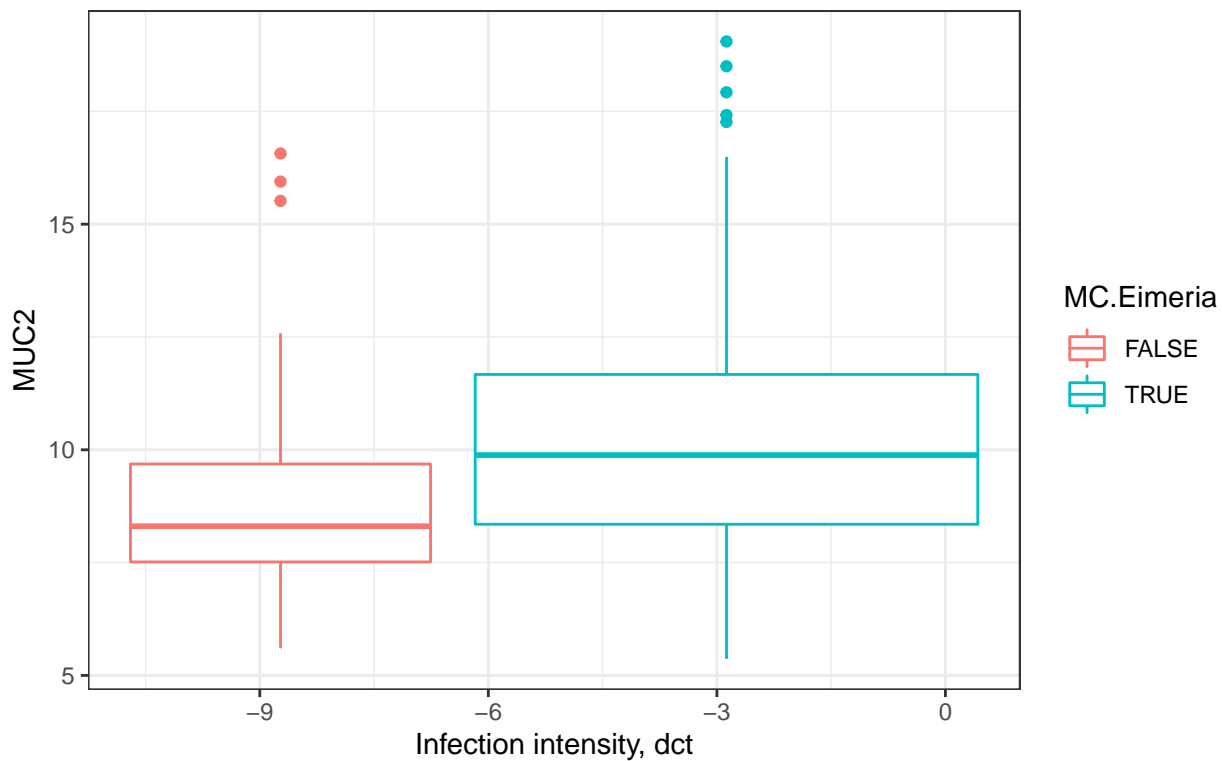## Warning: Removed 6 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MUC5AC,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

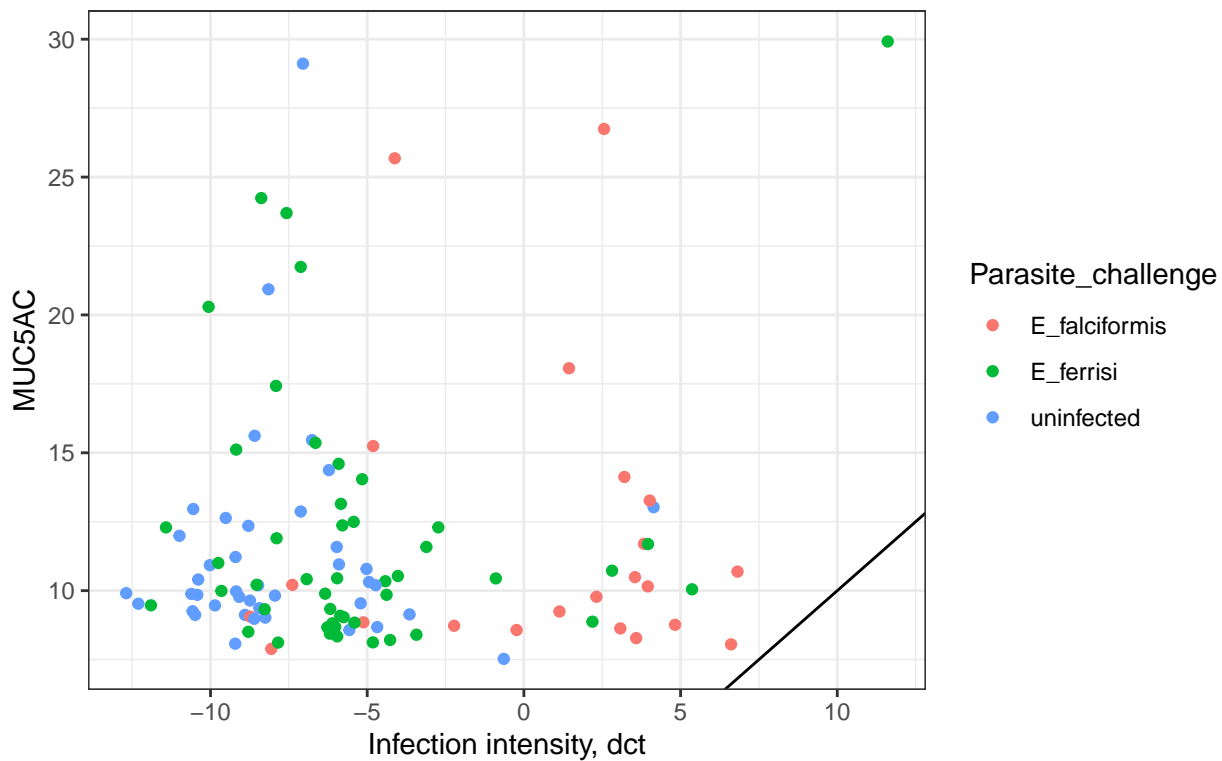# Immune gene expression against infection intensity in field samples

14. MYD88

```
ggplot(lab, aes(x = delta, y = MYD88, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

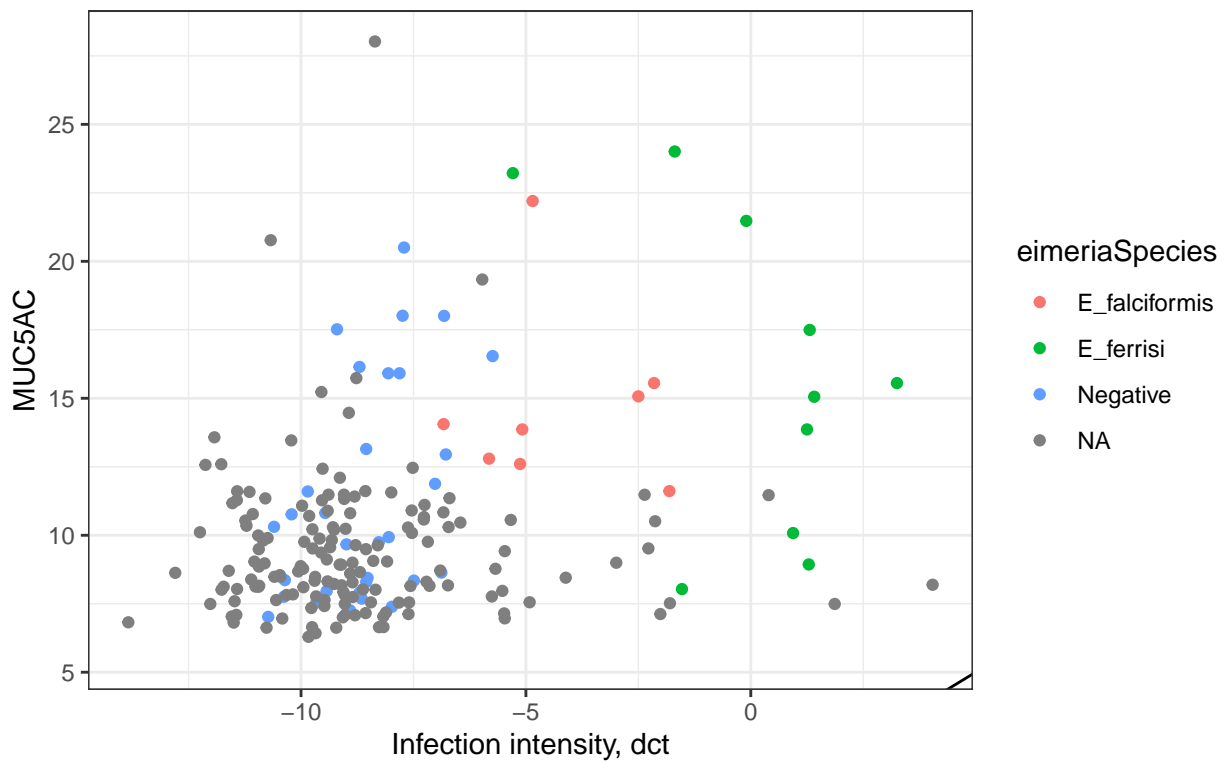## Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MYD88, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (geom_point).
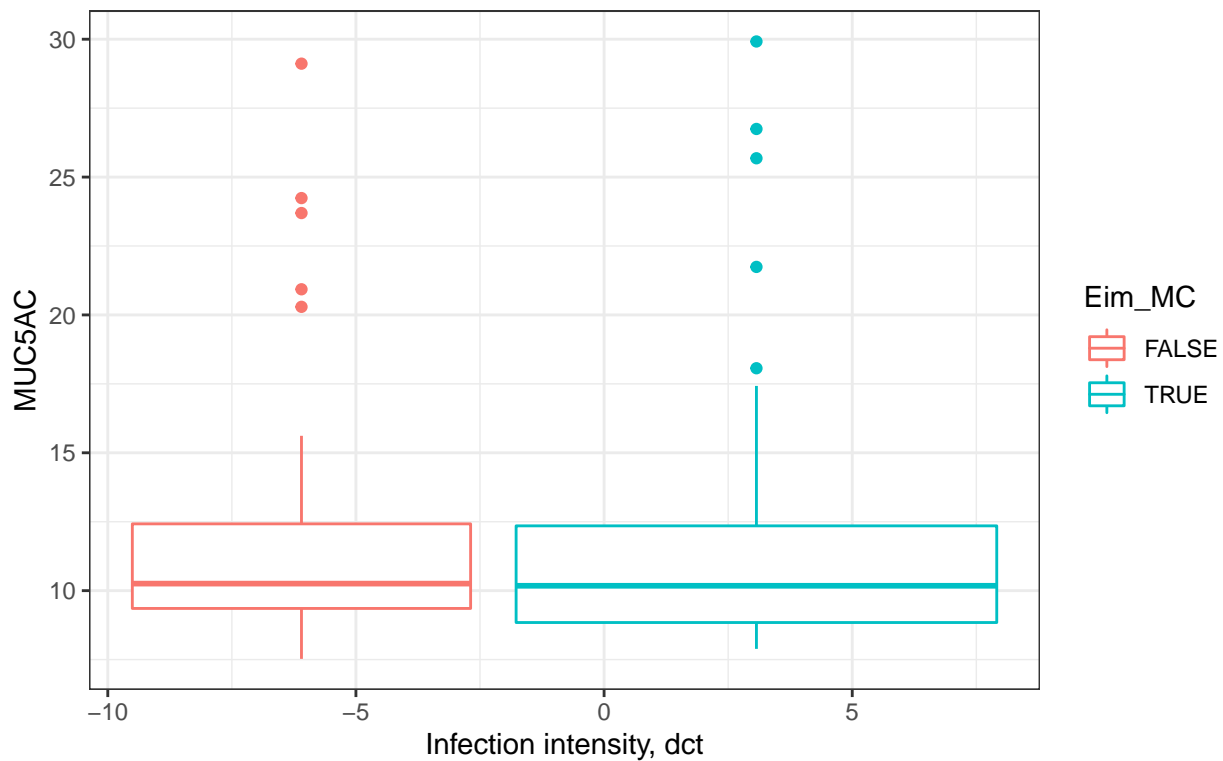
Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = MYD88, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```
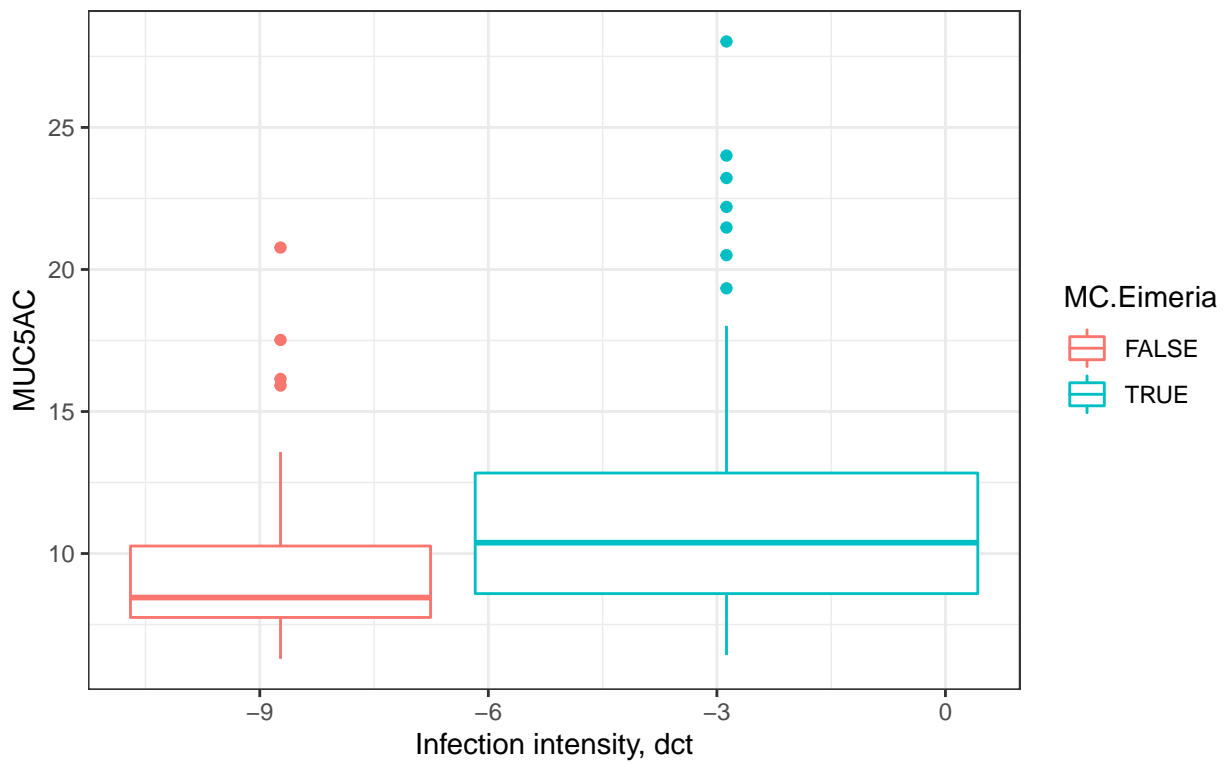
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = MYD88,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

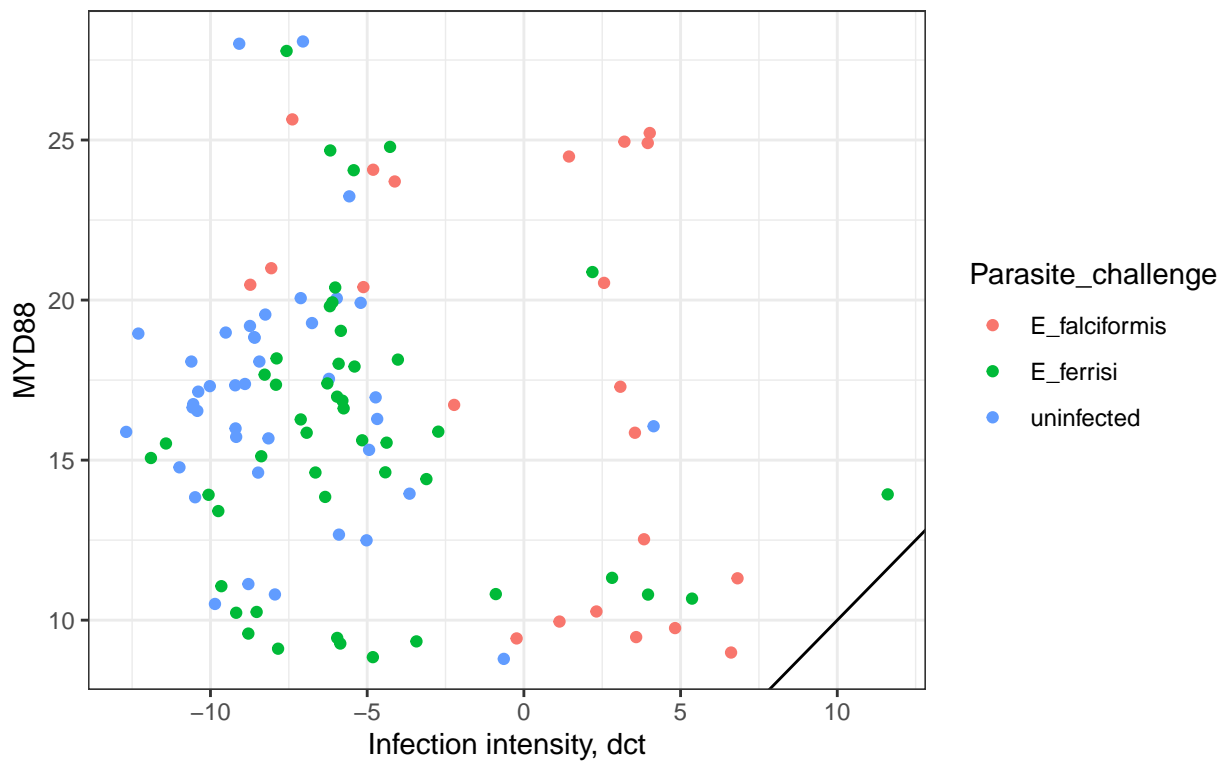Immune gene expression against infection intensity in field samples

NCR1

```
ggplot(lab, aes(x = delta, y = NCR1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

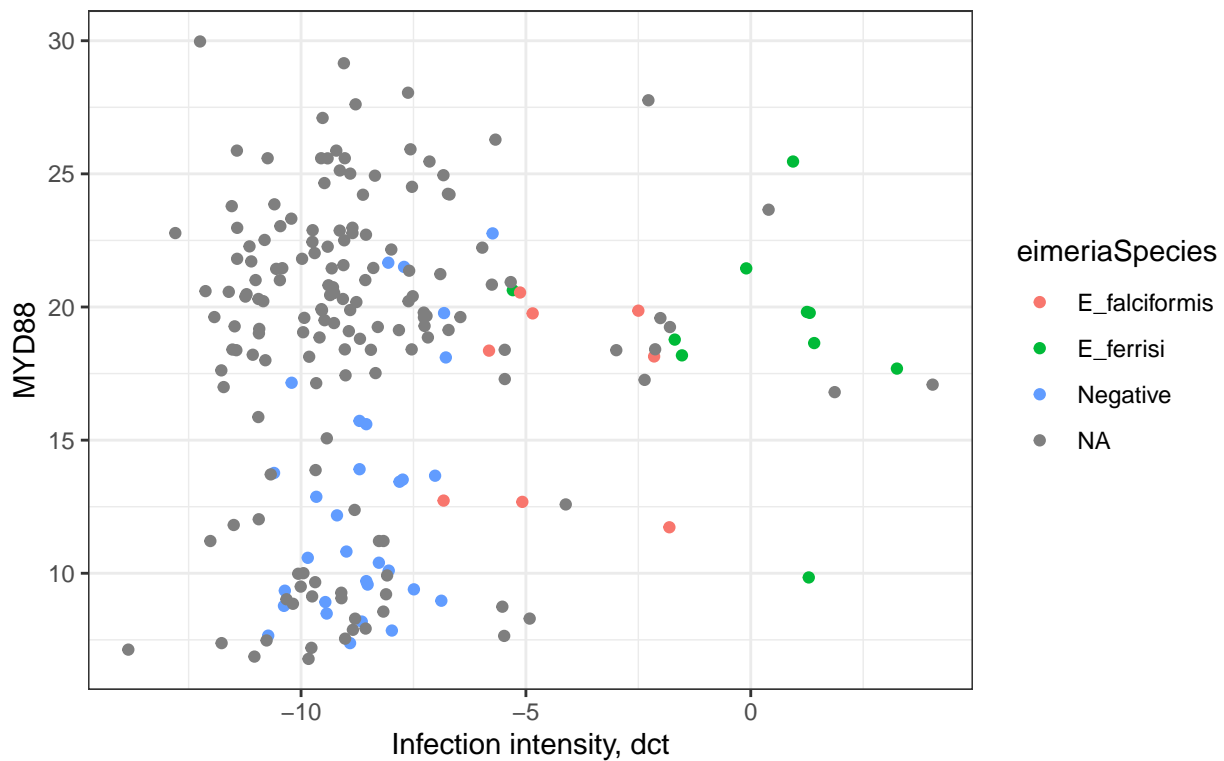## Warning: Removed 6 rows containing missing values (geom_point).

Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = NCR1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (geom_point).
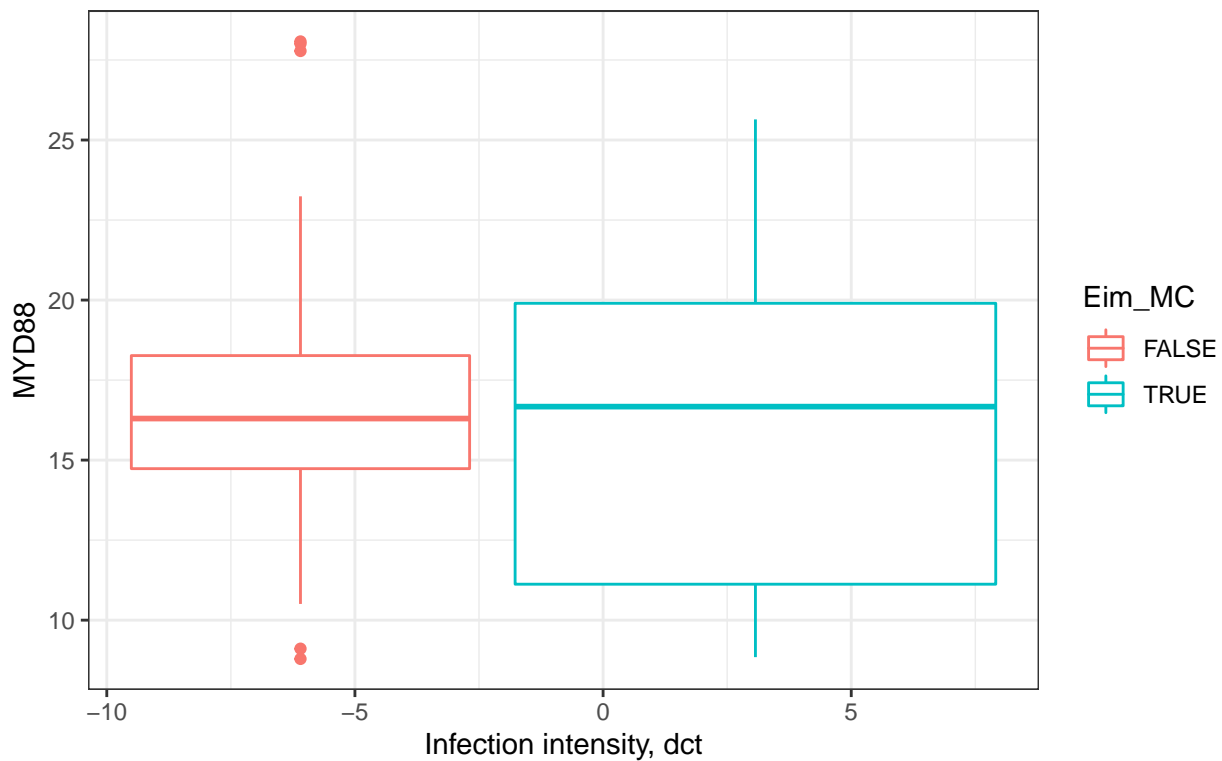
Immune gene expression against infection intensity in field samples

```
ggplot(lab, aes(x = delta, y = NCR1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```
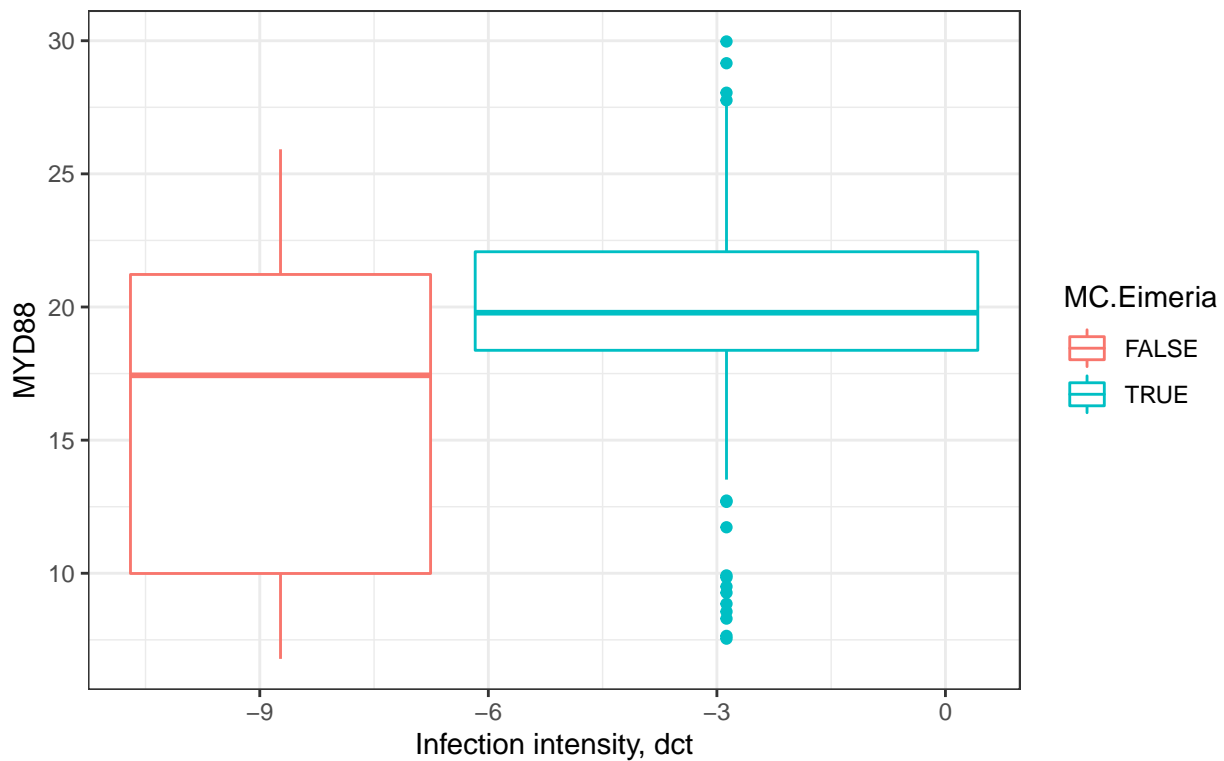
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = NCR1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

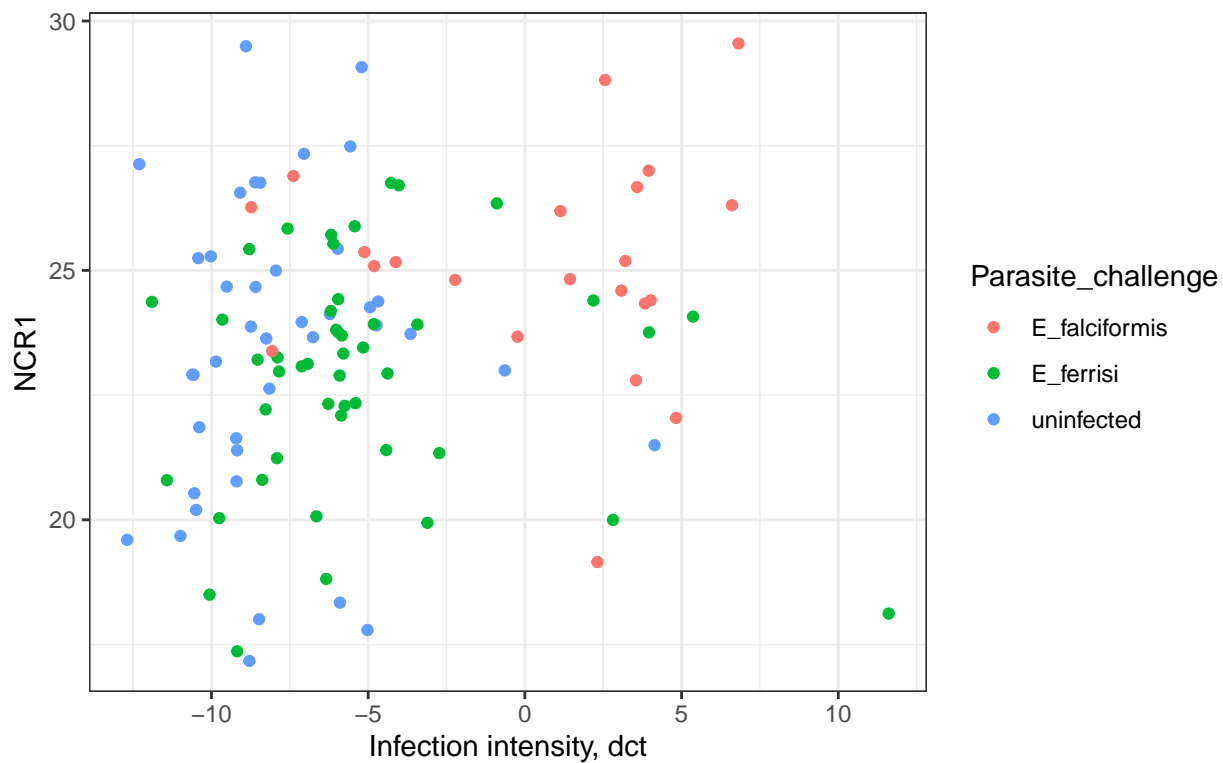# Immune gene expression against infection intensity in field samples

16. PRF1

```
ggplot(lab, aes(x = delta, y = PRF1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

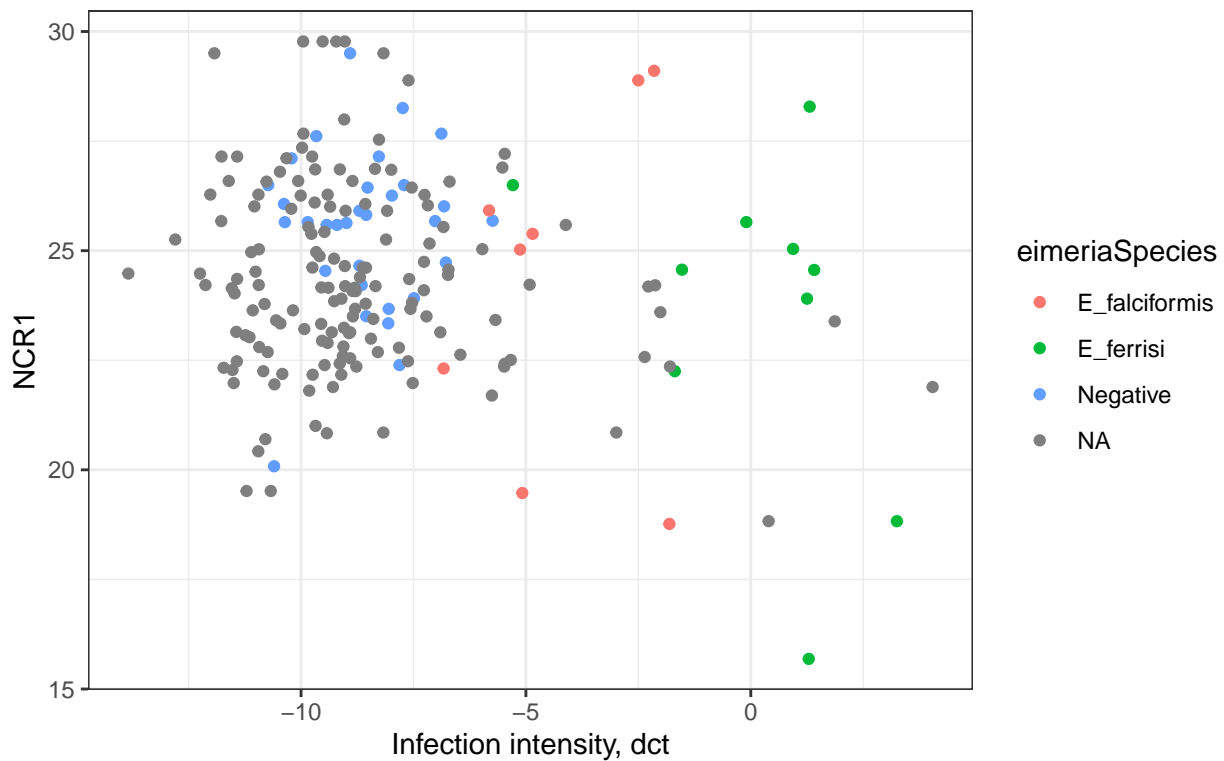Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = PRF1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```
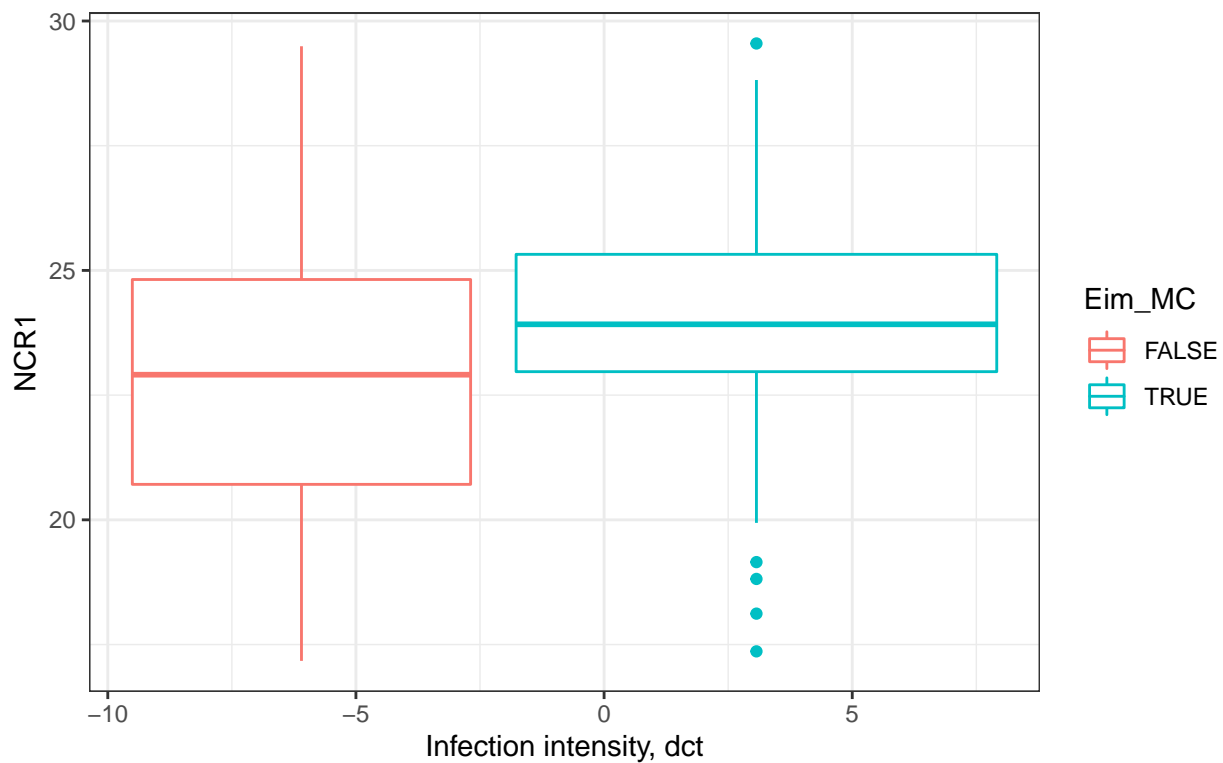
Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = PRF1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```

Immune gene expression against infection intensity in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = PRF1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

17. RETNLB

```
ggplot(lab, aes(x = delta, y = RETNLB, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
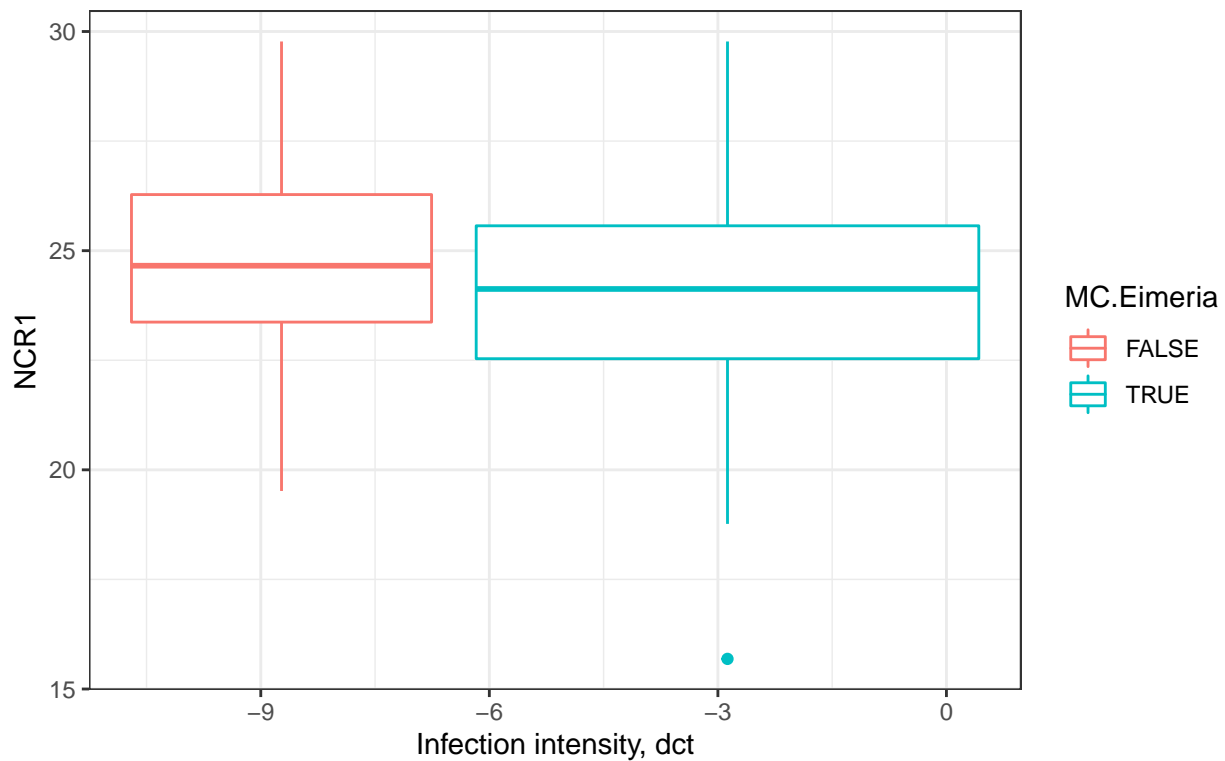
Immune gene expression against infection intensity
in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = RETNLB, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (geom_point).

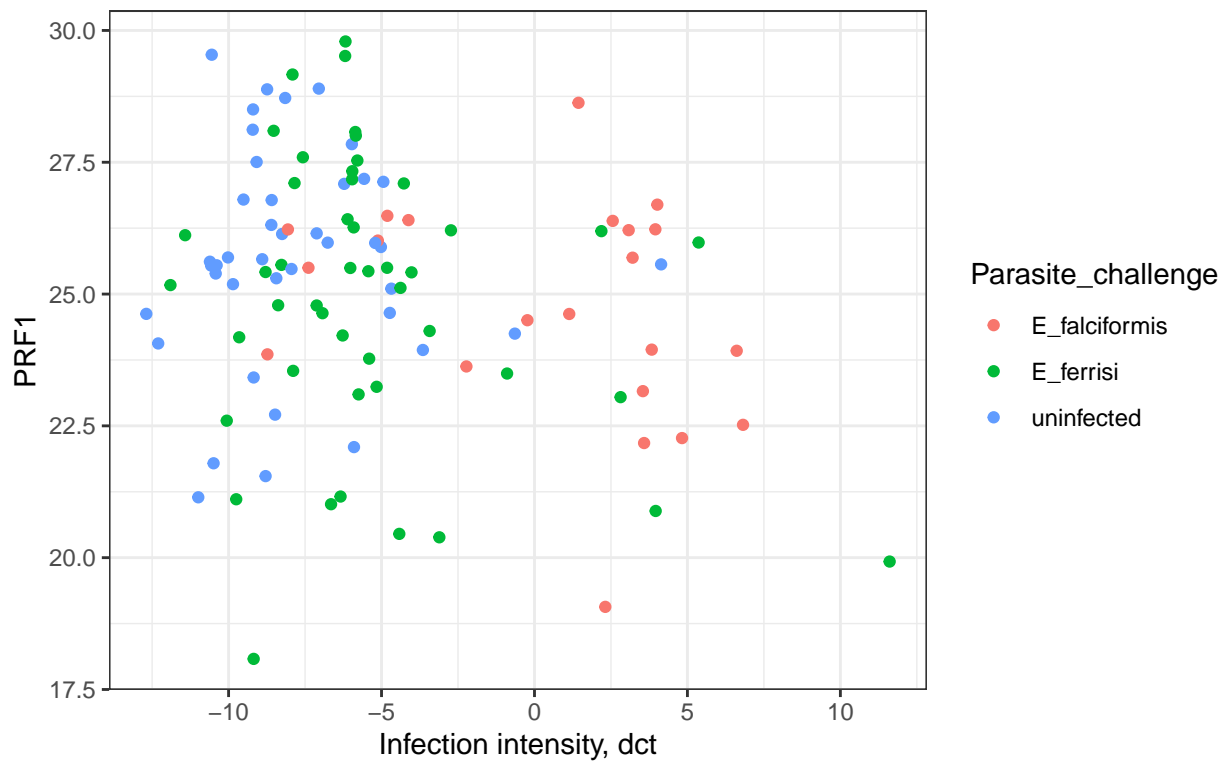## Immune gene expression against infection intensity in field samples


```
ggplot(lab, aes(x = delta, y = RETNLB, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```
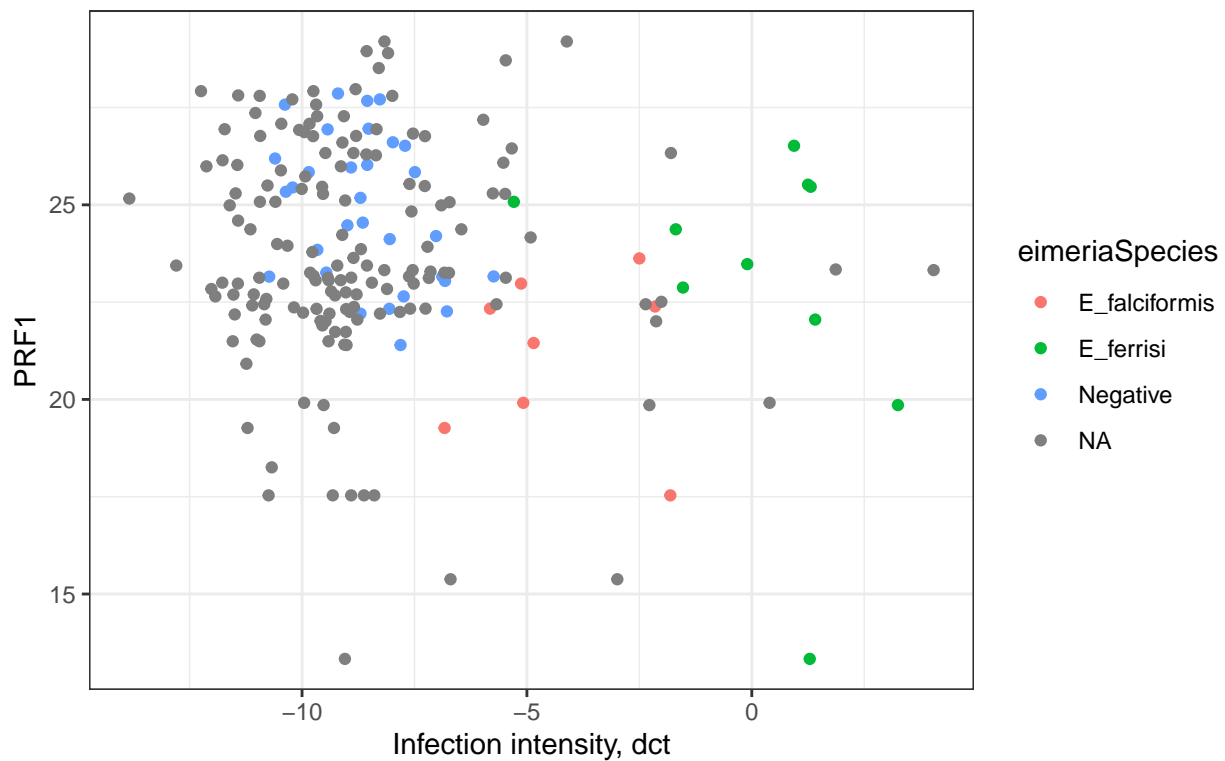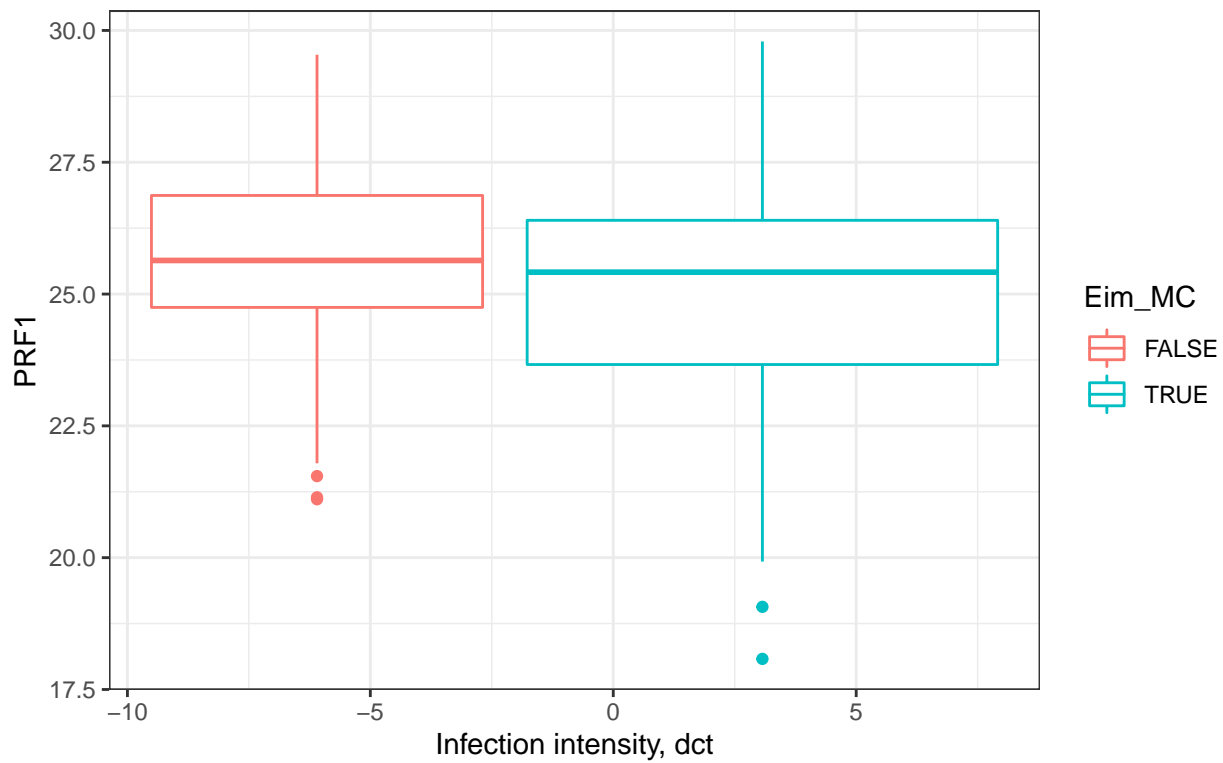
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = RETNLB,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (stat_boxplot).
```

Immune gene expression against infection intensity in field samples

18. SOCS1

```
ggplot(lab, aes(x = delta, y = SOCS1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
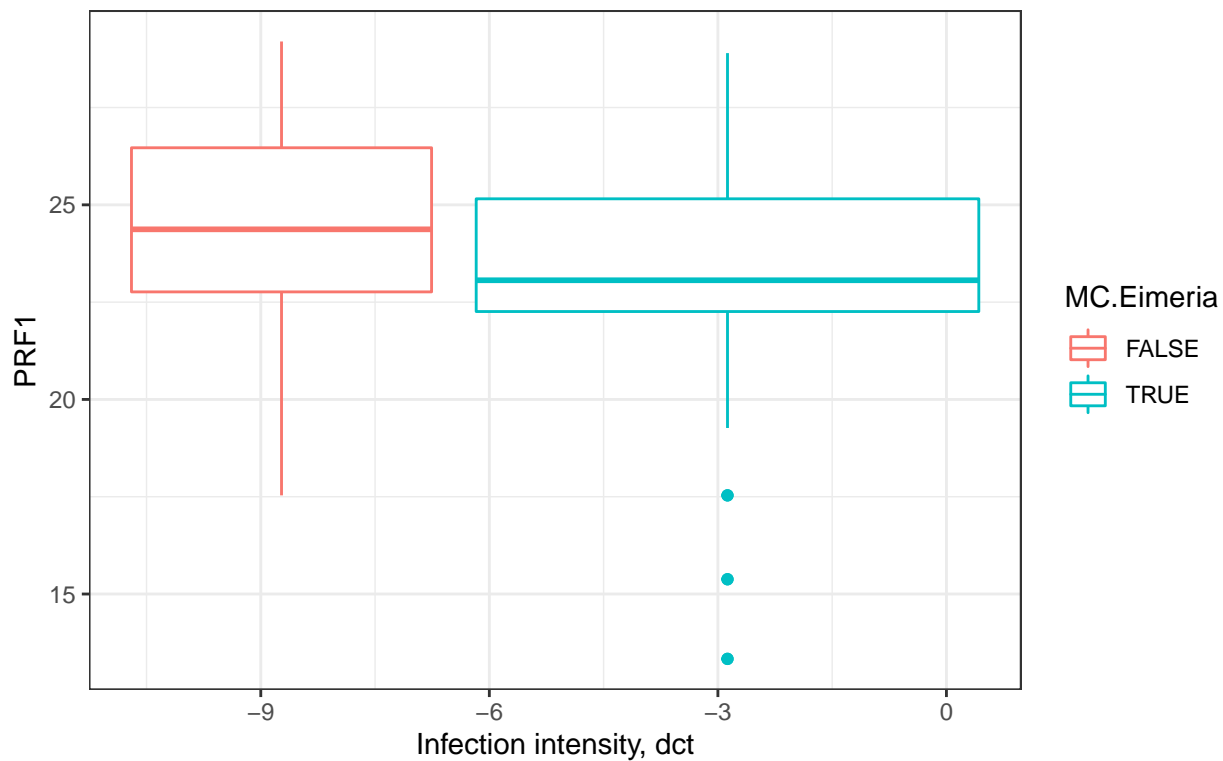
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = SOCS1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

```
## Warning: Removed 154 rows containing missing values (geom_point).
```

Immune gene expression against infection intensity in field samples

```
ggplot(lab, aes(x = delta, y = SOCS1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```
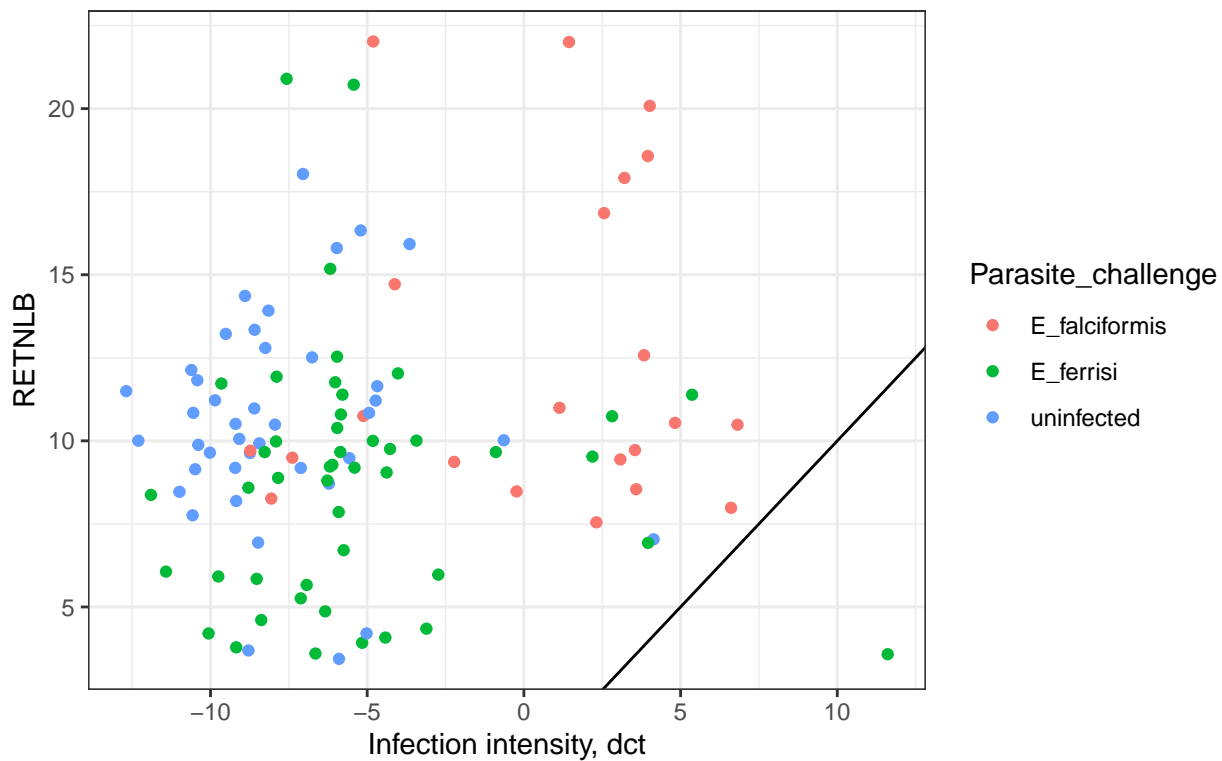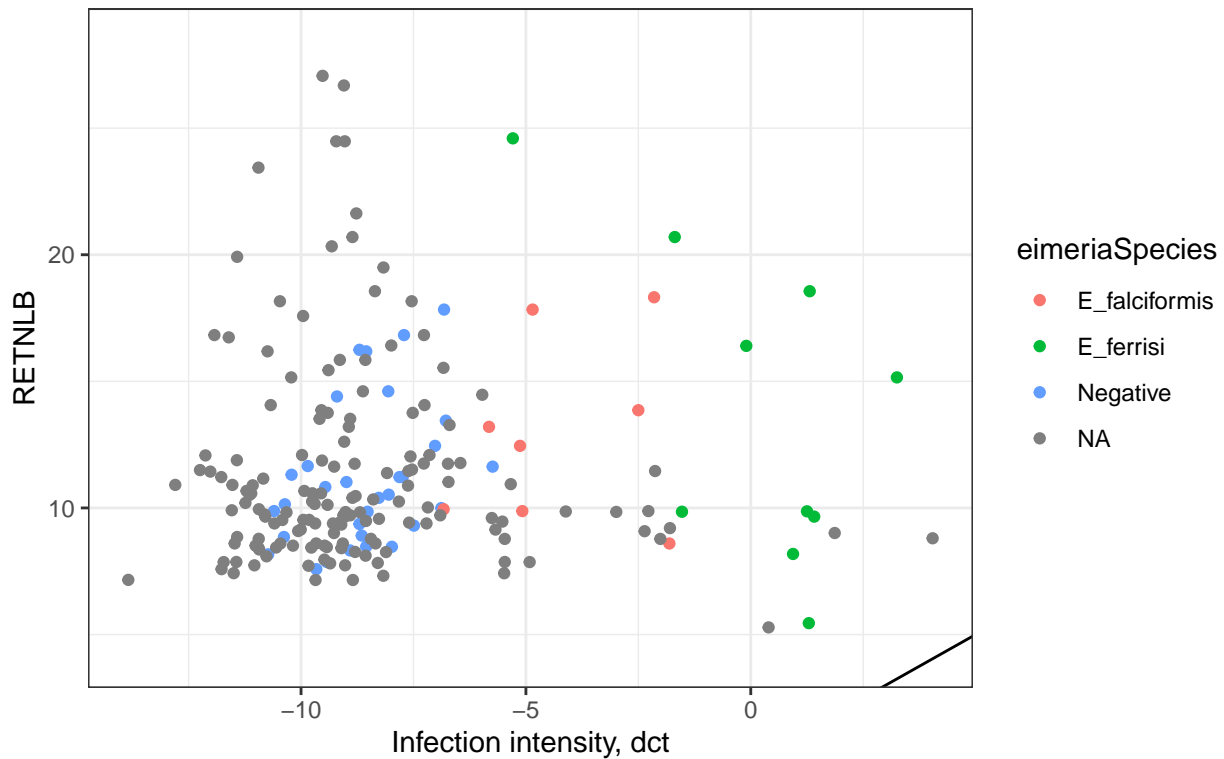
```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```

Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = SOCS1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).
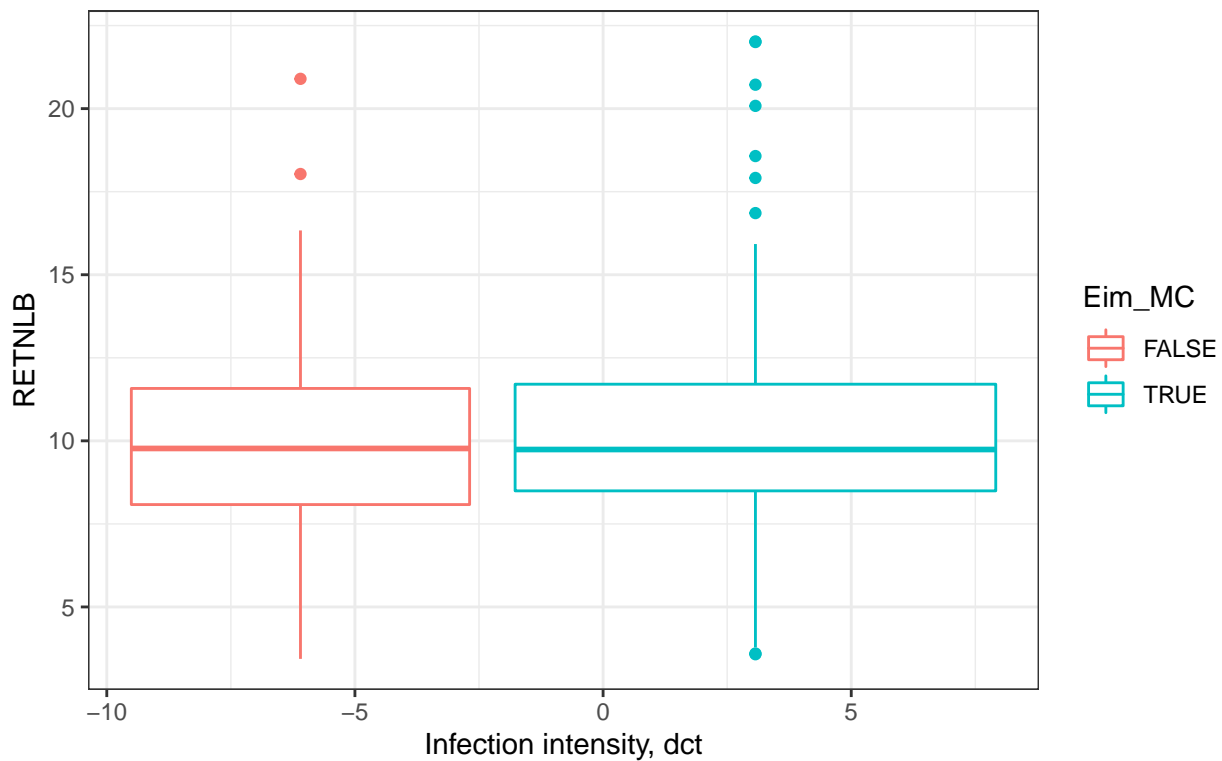
Immune gene expression against infection intensity in field samples

19. TICAM1

```
ggplot(lab, aes(x = delta, y = TICAM1, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```
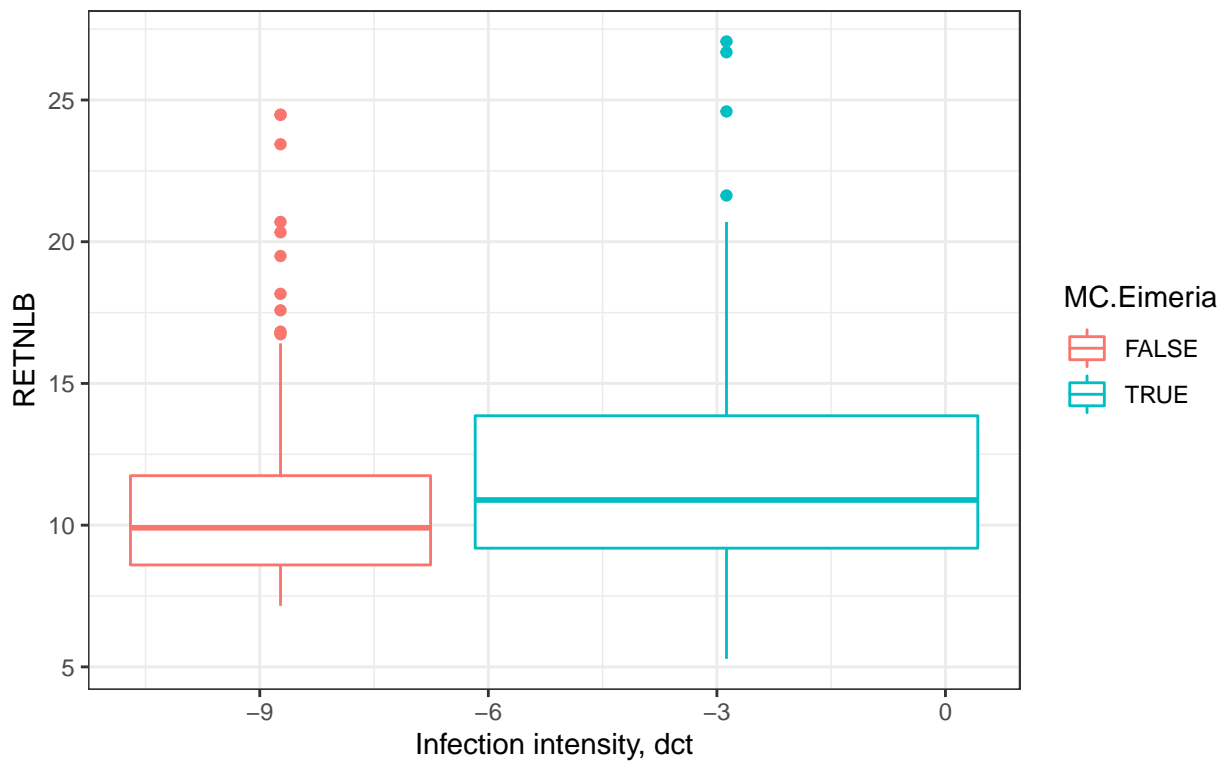
Immune gene expression against infection intensity
in lab infections

```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = TICAM1, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (geom_point).

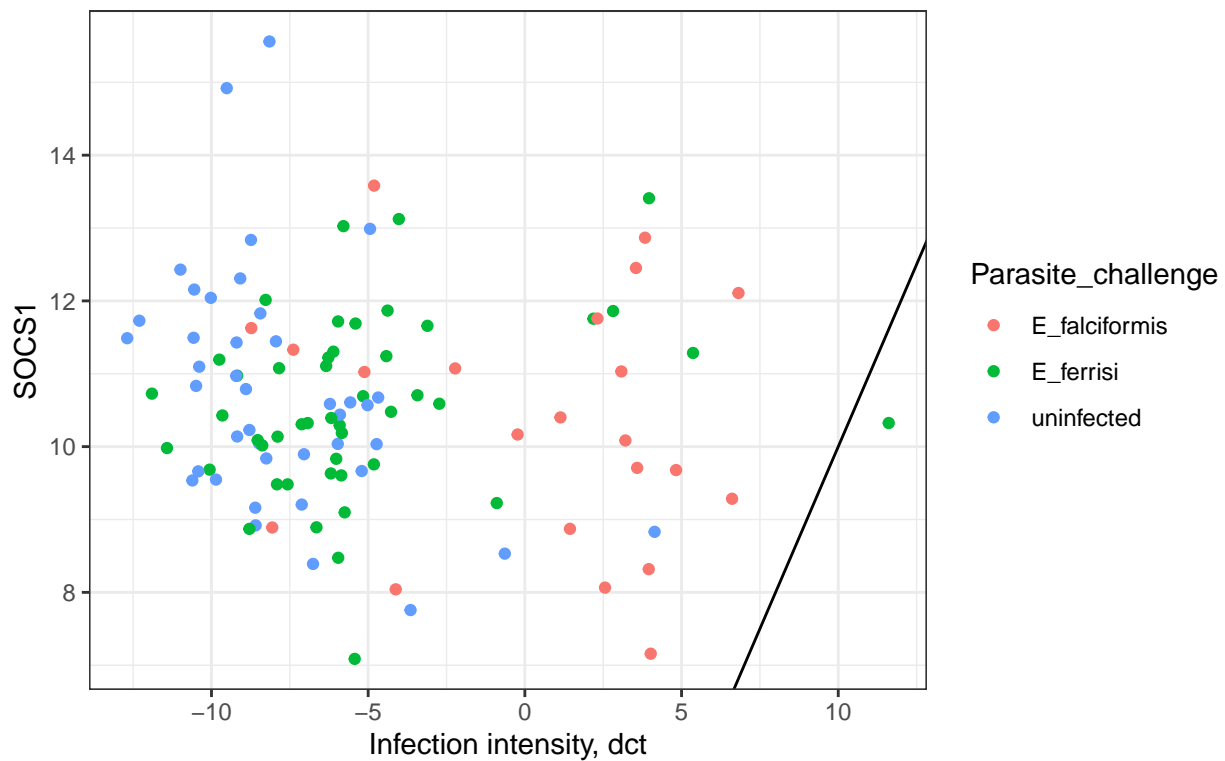Immune gene expression against infection intensity in field samples

```
ggplot(lab, aes(x = delta, y = TICAM1, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

Warning: Removed 6 rows containing missing values (stat_boxplot).
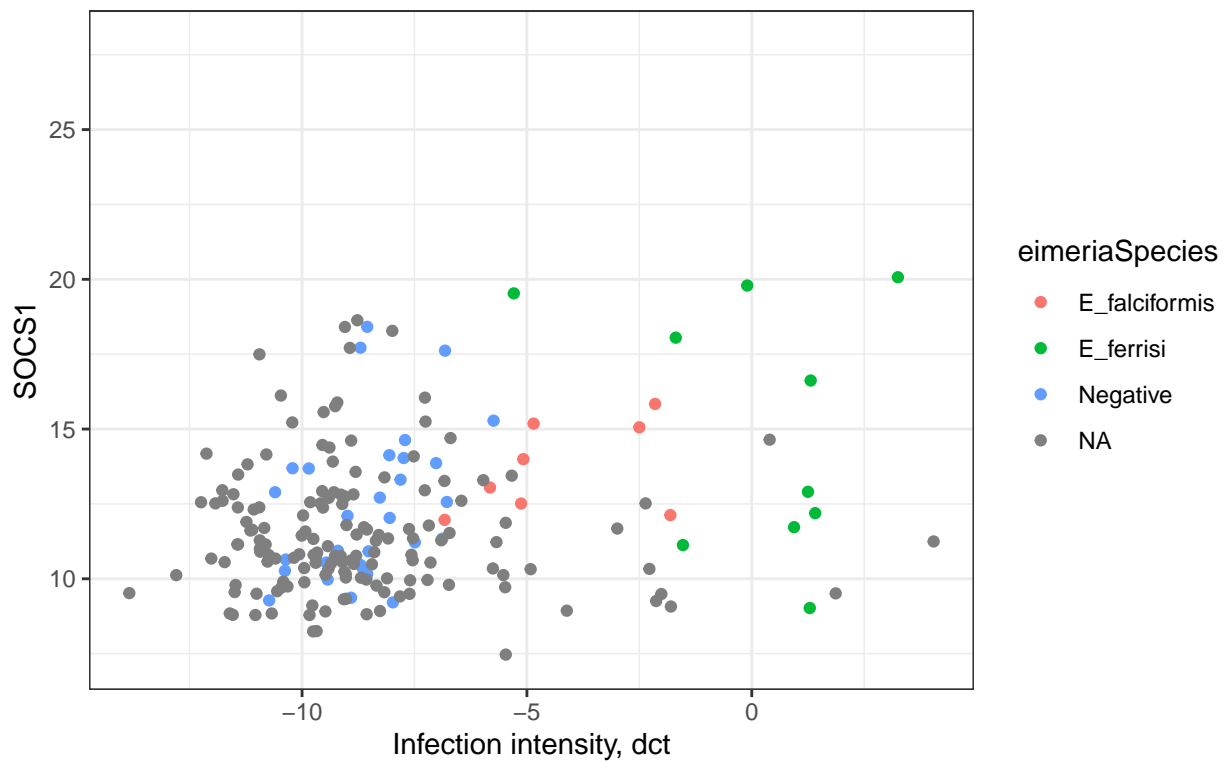
Immune gene expression against infection intensity in lab infections


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = TICAM1,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).

Immune gene expression against infection intensity in field samples

20. TNF

```
ggplot(lab, aes(x = delta, y = TNF, color = Parasite_challenge)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity
       in lab infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

Immune gene expression against infection intensity in lab infections
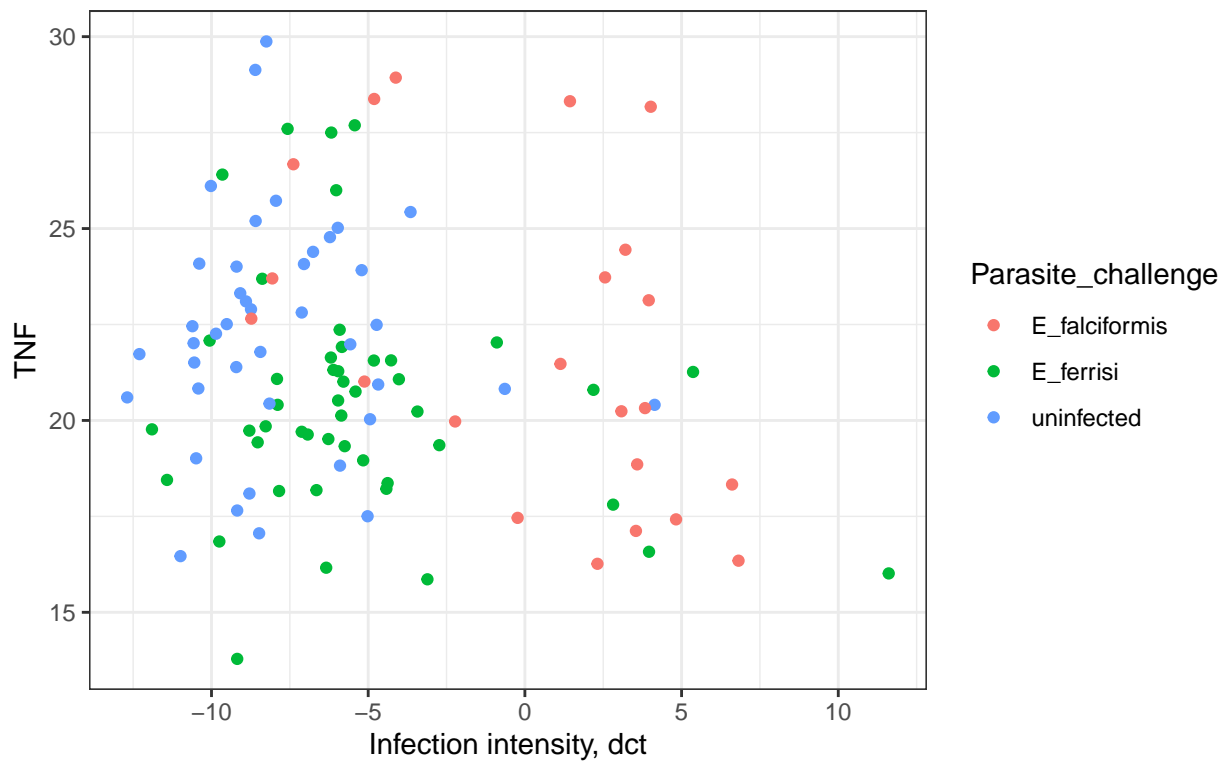
```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = TNF, color = eimeriaSpecies)) +
  geom_jitter() +
  geom_abline() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

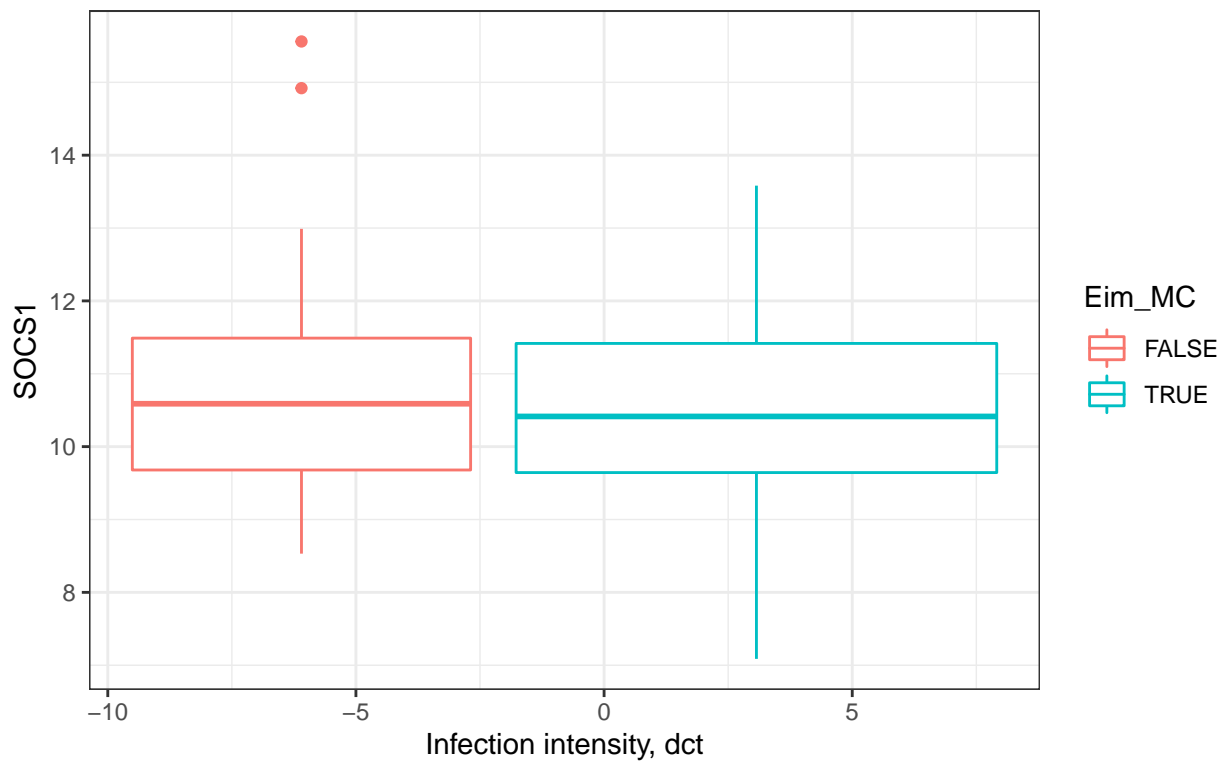## Warning: Removed 154 rows containing missing values (geom_point).

Immune gene expression against infection intensity in field samples
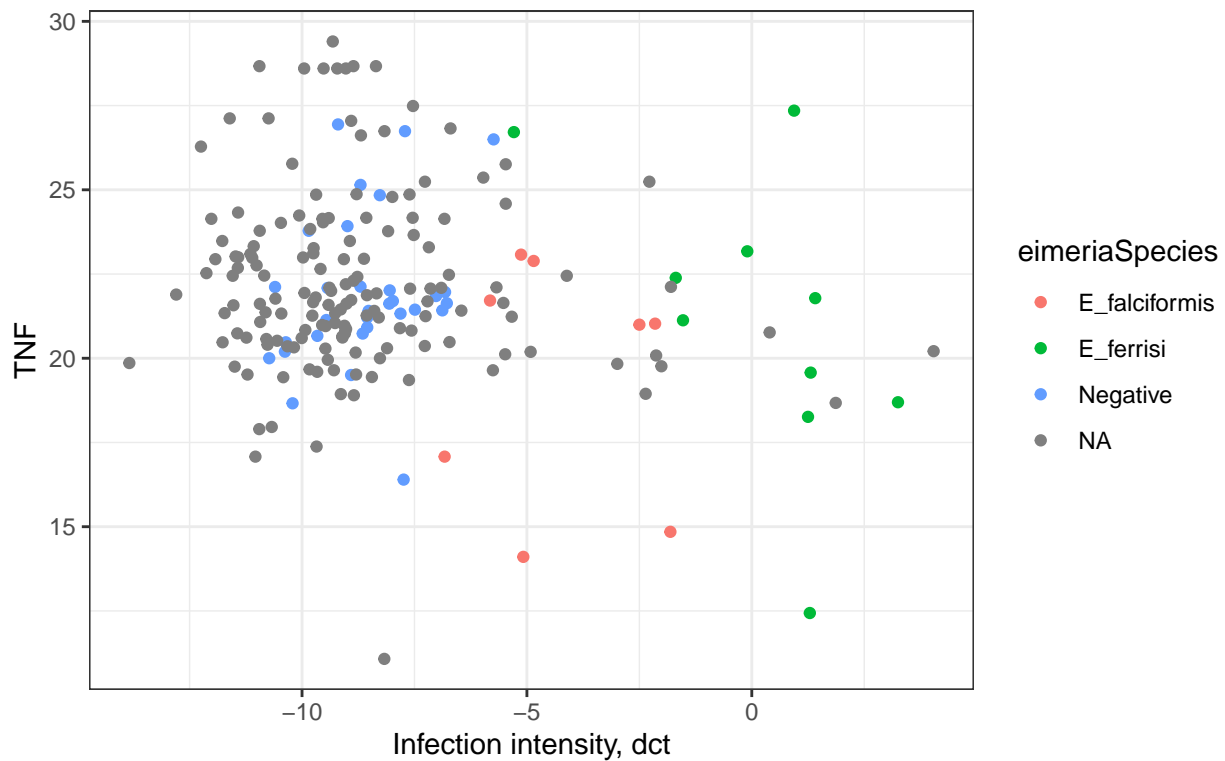
```
ggplot(lab, aes(x = delta, y = TNF, color = Eim_MC)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in lab
       infections") +
  theme_bw()
```

```
## Warning: Removed 6 rows containing missing values (stat_boxplot).
```
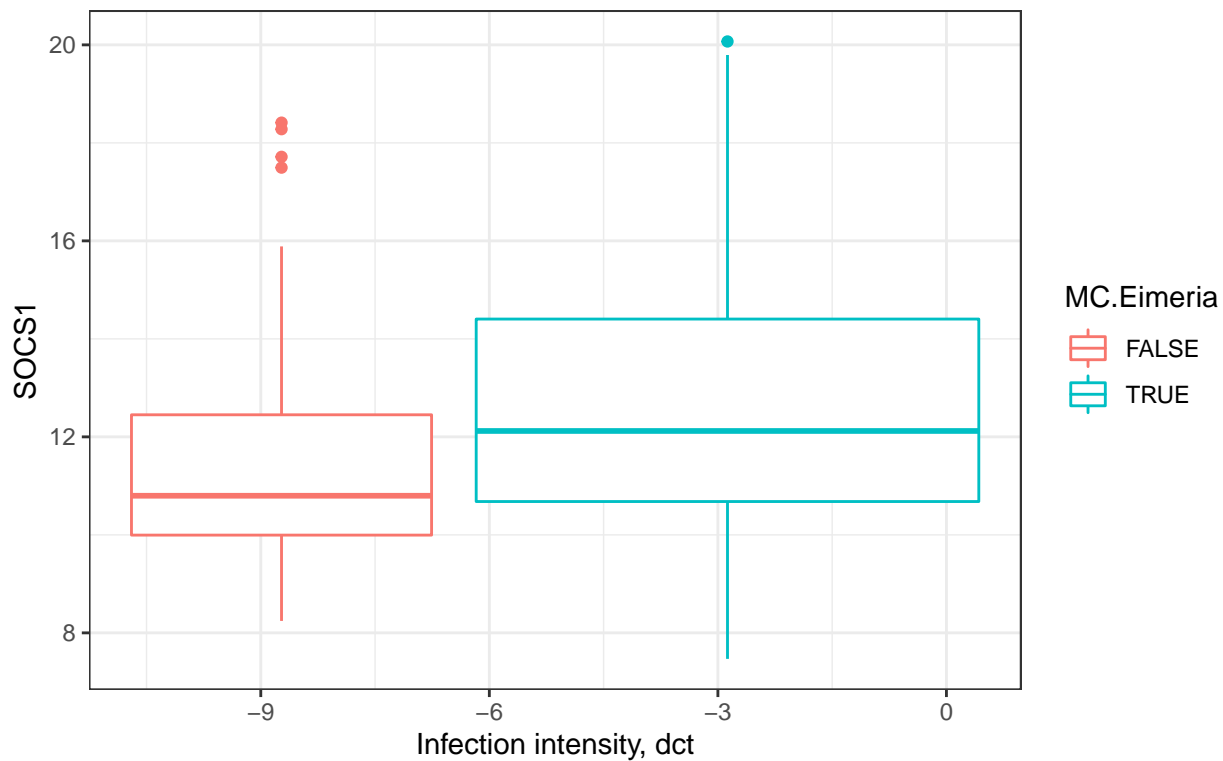
Immune gene expression against infection intensity in lab infections
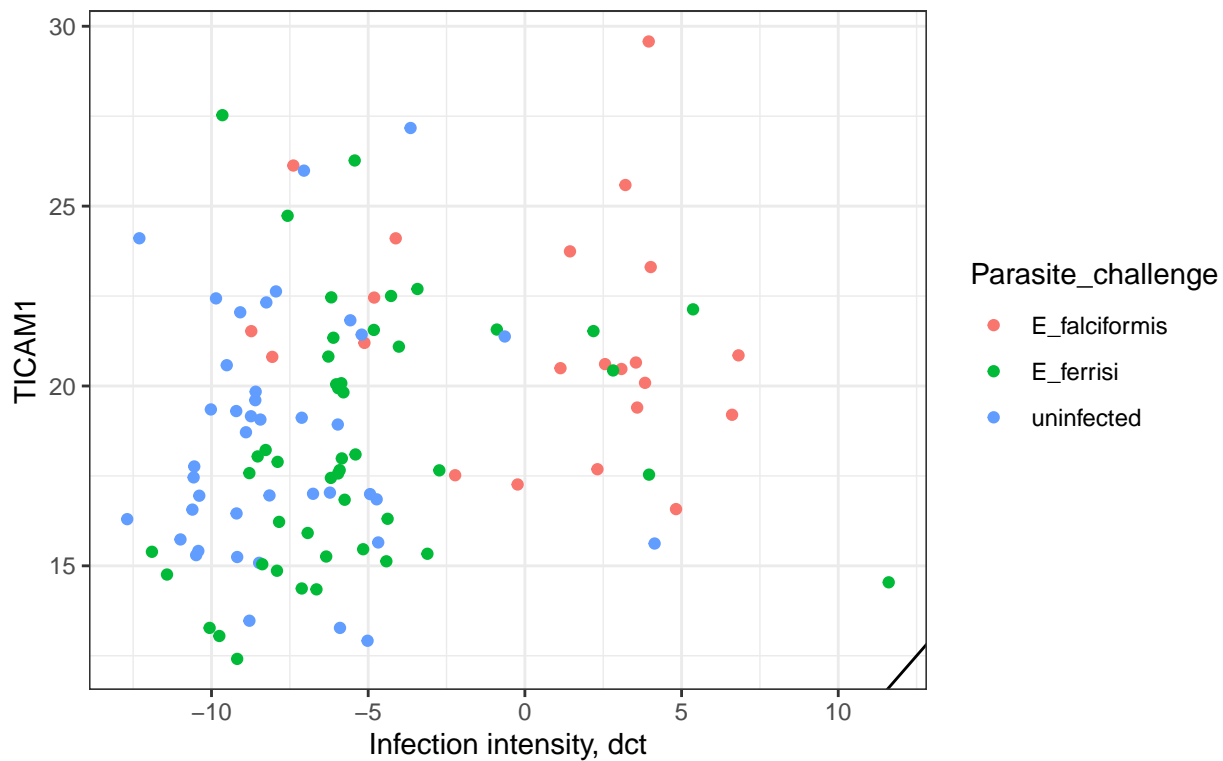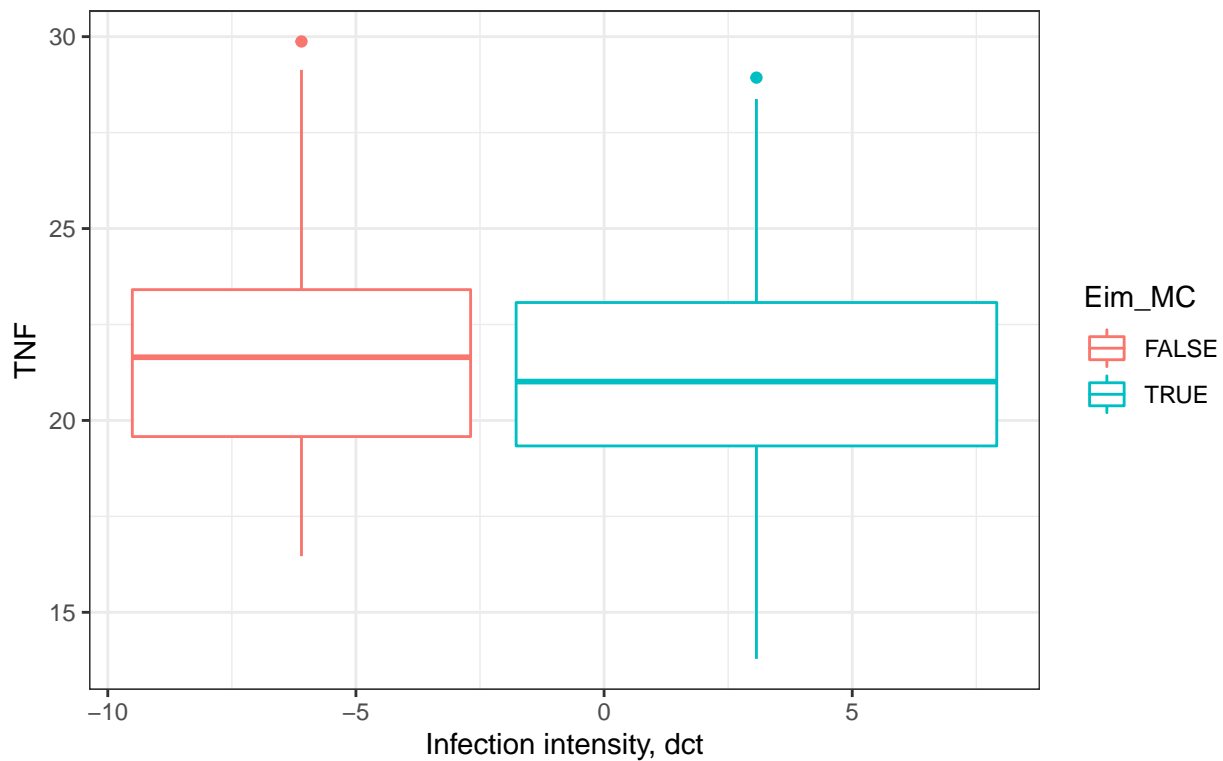


```
ggplot(field, aes(x = delta_ct_cewe_MminusE, y = TNF,
                  color = MC.Eimeria)) +
  geom_boxplot() +
  labs(x = "Infection intensity, dct",
       title ="Immune gene expression against infection intensity in field
       samples") +
  theme_bw()
```

Warning: Removed 154 rows containing missing values (stat_boxplot).
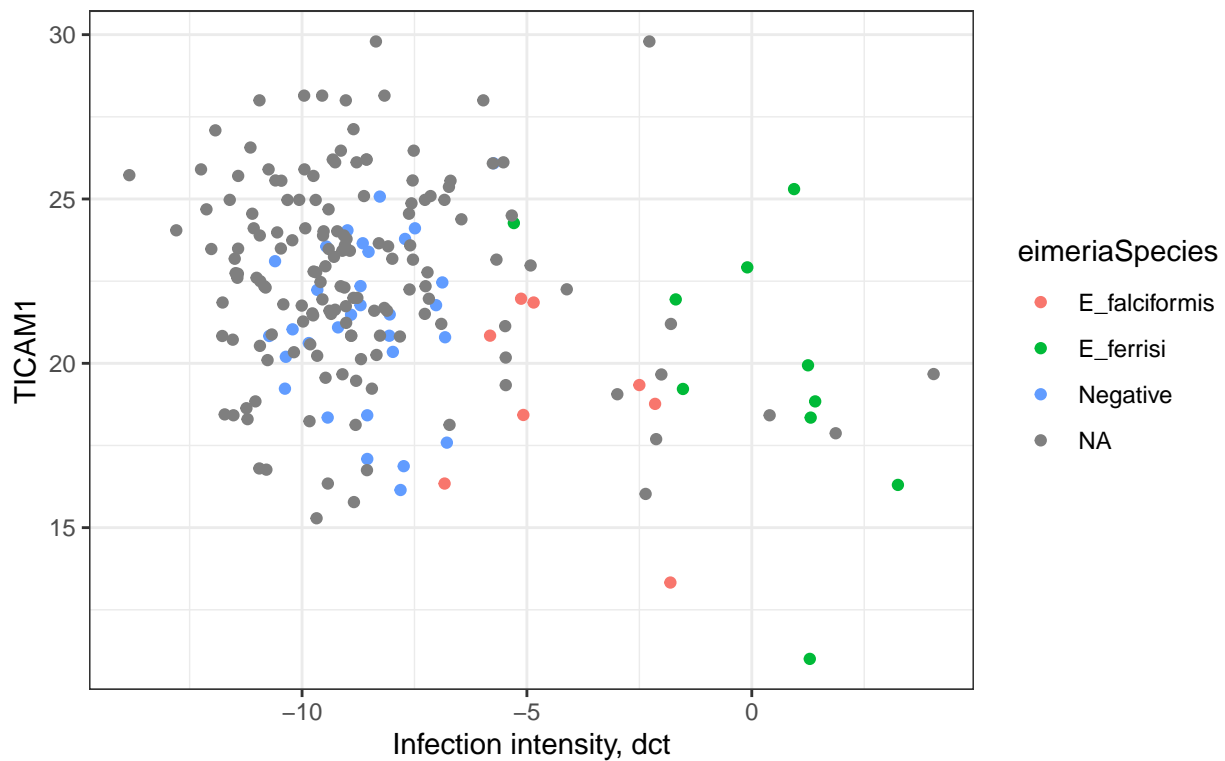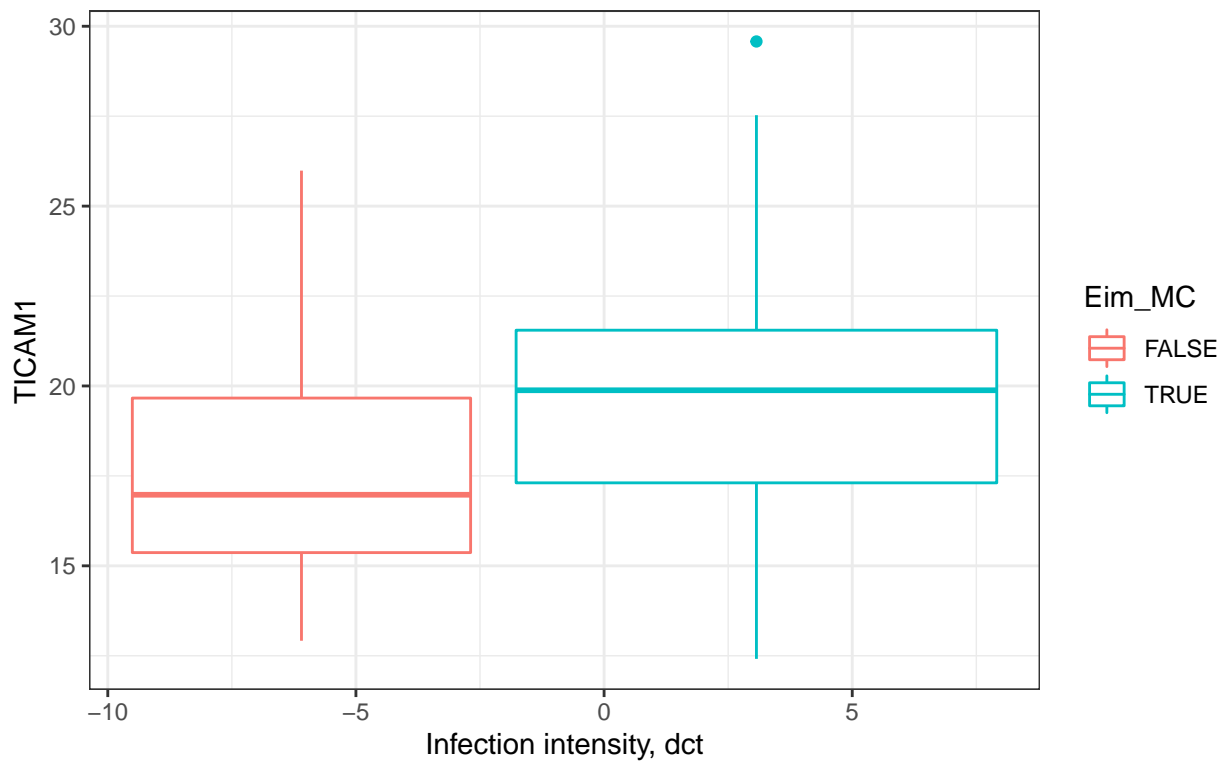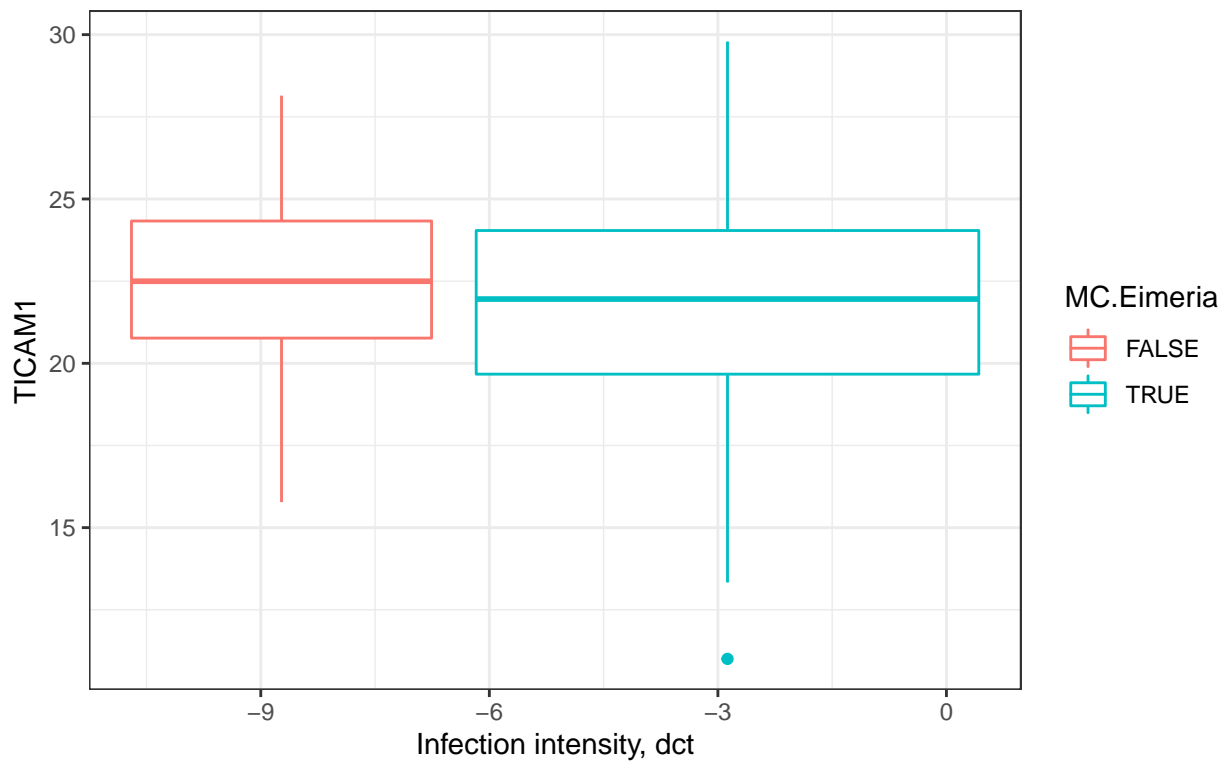
Immune gene expression against infection intensity in field samples
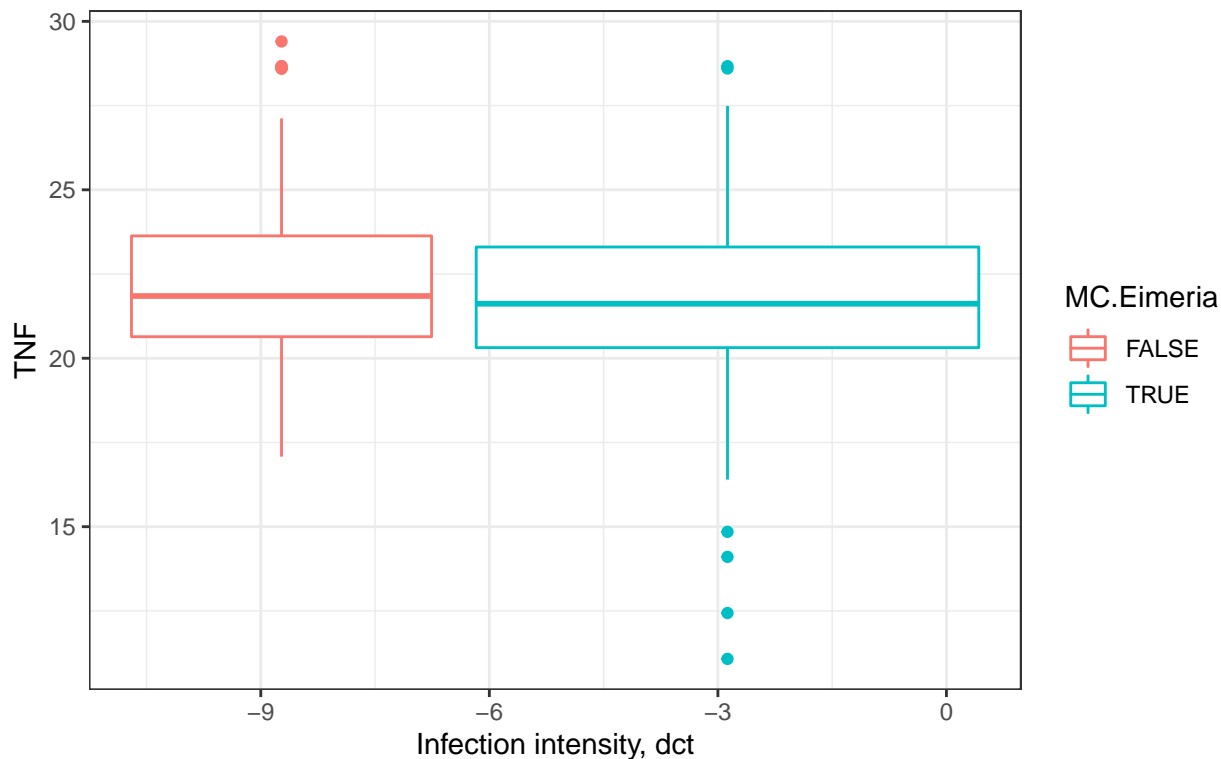
Compare the differences in uninfected and infected data and field / lab

Write a function which does a t test to quantify the differences of gene expression between uninfected and infected mice in the lab and field infections and also compares the uninfected lab and field and the infected lab and field

Function to print the p values for each t tet

```r
eimeria_t_test_pvalue <- function(x) {
 # select the uninfected (delta ct under -5 mice)
  lab_uni <- lab %>%
    dplyr::filter(Eim_MC == "FALSE")
# select the infected lab mice
  lab_inf <- lab %>%
    dplyr::filter(Eim_MC == "TRUE")
# select the uninfected field mice (delta ct under -5 mice)
  field_uni <- field %>%
    dplyr::filter(MC.Eimeria == "FALSE")
# select the infected lab mice
  field_inf <- field %>%
    dplyr::filter(MC.Eimeria == "TRUE")

lab.uni_lab.inf <- t.test(lab_uni %>% dplyr::select(all_of(x)),
                          lab_inf %>% dplyr::select(all_of(x)))$p.value

field.uni_field.inf <- t.test(field_uni %>% dplyr::select(all_of(x)),
```

```
                                field_inf %>% dplyr::select(all_of(x)))$p.value

lab.uni_field_uni <- t.test(lab_uni %>% dplyr::select(all_of(x)),
                             field_uni %>% dplyr::select(all_of(x)))$p.value


lab.inf_field.inf <- t.test(lab_inf %>% dplyr::select(all_of(x)),
                             field_inf %>% dplyr::select(all_of(x)))$p.value



A <- data.frame(lab.uni_lab.inf,field.uni_field.inf,lab.uni_field_uni,
                lab.inf_field.inf)

rownames(A) <- toString(x)

print(A)
}


P <- lapply(intersect(Genes_field, Genes_lab), eimeria_t_test_pvalue)
```

```
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IFNy   0.0008701234          0.02306666       0.009114009         0.4665444
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## CXCR3      0.7240836           0.9875716      6.892882e-23      1.609715e-30
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IL.6       0.4569897          0.06787592      0.0008034704        0.08853236
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IL.10      0.6664722           0.6846399       0.007287372      0.0004755641
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IL.13     0.01136685        8.790546e-05      3.030428e-06      4.358343e-16
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IL1RN      0.7694255          0.08933739      2.494644e-07      0.0002133729
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## CASP1      0.9965703          0.05714562        0.08776055         0.9422483
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## CXCL9   3.645686e-05          0.03663115       0.001032312      5.396998e-10
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IDO1     0.01723246           0.1606432      3.080484e-05          0.601396
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## IRGM1    0.03189147         0.008218303        0.01225175      2.273495e-10
##     lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## MPO       0.6247867          0.01078007      3.373386e-07       0.004021206
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## MUC2       0.4783091         1.14159e-05         0.5692909      3.134409e-06
##       lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## MUC5AC       0.8906956        9.076836e-06       0.001266663         0.9759575
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## MYD88      0.6477168        1.543588e-05         0.7718497      6.623496e-05
##     lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## NCR1      0.1069354          0.01499511      0.0004910738         0.6735697
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## PRF1       0.1370779         0.004534779       0.001882949      5.780491e-05
##       lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## RETNLB       0.6017666         0.08135791        0.09105424        0.01363267
##      lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## SOCS1      0.2327638        0.0001628878        0.07303862      1.128447e-09
```

```
##          lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## TICAM1        0.04802866           0.2056721      6.600607e-10        8.18763e-05
##       lab.uni_lab.inf field.uni_field.inf lab.uni_field_uni lab.inf_field.inf
## TNF        0.4766501           0.1809073         0.2633122          0.394101
```

```r
t.test.p.value <- reduce(P, rbind)

t.test.p.value <- tibble::rownames_to_column(t.test.p.value, "Genes")

write.csv(t.test.p.value,
          "output_data/gene_expression/data_products/t.test.p.values_gene_exp.csv",
          row.names = FALSE)
```

Which genes are significant to predict weight loss?

```r
Lab_wl <- lm(formula = max_WL ~ IFNy + CXCR3 + IL.6 + IL.10 + IL.13 + IL1RN +
                CXCR3 + CASP1 + CXCL9 + IDO1 + IRGM1 + MPO + MUC2 + MUC5AC +
                MYD88 + NCR1 + PRF1 + RETNLB + SOCS1 + TICAM1 + TNF, data = lab)

summary(Lab_wl)
```

```
##
## Call:
## lm(formula = max_WL ~ IFNy + CXCR3 + IL.6 + IL.10 + IL.13 + IL1RN +
##     CXCR3 + CASP1 + CXCL9 + IDO1 + IRGM1 + MPO + MUC2 + MUC5AC +
##     MYD88 + NCR1 + PRF1 + RETNLB + SOCS1 + TICAM1 + TNF, data = lab)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6176 -2.4799  0.1734  1.8098  9.5301
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.000006  15.405379  -0.195   0.8466
## IFNy         0.322858   0.428596   0.753   0.4557
## CXCR3       -0.092789   0.140074  -0.662   0.5115
## IL.6        -0.037201   0.319540  -0.116   0.9079
## IL.10       -0.074093   0.412013  -0.180   0.8582
## IL.13        1.304887   0.948274   1.376   0.1765
## IL1RN       -0.797456   0.571481  -1.395   0.1706
## CASP1       -0.616740   0.499440  -1.235   0.2241
## CXCL9       -0.241405   0.574569  -0.420   0.6766
## IDO1        -0.878372   0.464389  -1.891   0.0658 .
## IRGM1        0.175299   1.221502   0.144   0.8866
## MPO          0.371152   0.359642   1.032   0.3083
## MUC2        -0.074745   0.671834  -0.111   0.9120
## MUC5AC      -0.179853   0.316862  -0.568   0.5735
## MYD88       -0.547750   0.695541  -0.788   0.4356
## NCR1        -0.227017   0.452027  -0.502   0.6183
## PRF1         0.510498   0.381384   1.339   0.1883
## RETNLB      -0.003365   0.350031  -0.010   0.9924
## SOCS1       -0.174075   1.442321  -0.121   0.9045
## TICAM1       0.861394   0.749231   1.150   0.2571
## TNF          0.221457   0.475055   0.466   0.6436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.214 on 40 degrees of freedom
```

```
##    (55 observations deleted due to missingness)
## Multiple R-squared:  0.4005, Adjusted R-squared:  0.1008
## F-statistic: 1.336 on 20 and 40 DF,  p-value: 0.213
```

```
AIC(Lab_wl)
```

```
## [1] 366.853
```

```
Lab_MC <- lm(formula = Eim_MC ~ IFNy + CXCR3 + IL.6 + IL.10 + IL.13 + IL1RN +
                CXCR3 + CASP1 + CXCL9 + IDO1 + IRGM1 + MPO + MUC2 + MUC5AC +
                MYD88 + NCR1 + PRF1 + RETNLB + SOCS1 + TICAM1 + TNF, data = lab)
```

```
summary(Lab_MC)
```

```
##
## Call:
## lm(formula = Eim_MC ~ IFNy + CXCR3 + IL.6 + IL.10 + IL.13 + IL1RN +
##     CXCR3 + CASP1 + CXCL9 + IDO1 + IRGM1 + MPO + MUC2 + MUC5AC +
##     MYD88 + NCR1 + PRF1 + RETNLB + SOCS1 + TICAM1 + TNF, data = lab)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96381 -0.21804  0.03771  0.22030  0.87795
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.5822570  1.5120281  -0.385   0.7022
## IFNy        -0.0009029  0.0420665  -0.021   0.9830
## CXCR3       -0.0103727  0.0137482  -0.754   0.4550
## IL.6         0.0081549  0.0313627   0.260   0.7962
## IL.10        0.0193551  0.0404388   0.479   0.6348
## IL.13        0.0725771  0.0930725   0.780   0.4401
## IL1RN        0.0319357  0.0560905   0.569   0.5723
## CASP1        0.0332448  0.0490198   0.678   0.5016
## CXCL9       -0.0792809  0.0563935  -1.406   0.1675
## IDO1        -0.0899176  0.0455795  -1.973   0.0555 .
## IRGM1       -0.1607594  0.1198897  -1.341   0.1875
## MPO         -0.0072732  0.0352986  -0.206   0.8378
## MUC2        -0.0175636  0.0659401  -0.266   0.7913
## MUC5AC      -0.0113556  0.0310998  -0.365   0.7169
## MYD88        0.0886323  0.0682669   1.298   0.2016
## NCR1        -0.0240597  0.0443661  -0.542   0.5906
## PRF1         0.0210632  0.0374326   0.563   0.5768
## RETNLB       0.0075148  0.0343553   0.219   0.8280
## SOCS1        0.0706436  0.1415629   0.499   0.6205
## TICAM1       0.0674897  0.0735366   0.918   0.3642
## TNF         -0.0476275  0.0466263  -1.021   0.3132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4136 on 40 degrees of freedom
##    (55 observations deleted due to missingness)
## Multiple R-squared:  0.5502, Adjusted R-squared:  0.3253
## F-statistic: 2.447 on 20 and 40 DF,  p-value: 0.007899
```

```
AIC(Lab_MC)
```

```
## [1] 83.65866
```

Results

1. Significant difference between means of lab uni and lab inf:

- IFNy
- IL.13
- PRF1
- TICAM1

2. Significant difference between means of field uni and field inf:

- IFNy
- IL.6
- IL.13
- IL1RN
- CXCL9
- MPO
- MUC5AC
- PRF1
- TICAM1
- TNF

3. Significant difference between means of lab uni and field uni: IFNy CXCR3 IL.6 IL.10 IL.13 IL1RN CXCL9 IDO1 IRGM1 MPO MUC2 MUC5AC NCR1 PRF1 RETNLB SOCS1 TICAM1

4. No Significant difference between means of lab uni and field uni:

- CASP1
- MYD88 -TNF

5. Significant difference between means of lab inf and field inf: IFNy CXCR3 IL.13 IRGM1 MUC2 MYD88 PRF1 SOCS1

6. No Significant difference between means of lab inf and field inf: IL.6 IL.10 IL1RN CASP1 CXCL9 IDO1 MPO MUC5AC NCR1 RETNLB TICAM1 TNF

https://docs.google.com/spreadsheets/d/1JWjsyzlQxWlSttZ5KwxYSsZapjZCgMCO1OcF0Yz9_ZQ/edit?usp=sharing

It is time to apply the package of Alice Balard et al. on our predictions!

Let's see if we indeed have differences across the hybrid index with our predicted weight loss.

Install the package

```
##
## * checking for file '/tmp/RtmpzAleVu/remotes2e5e9940fa02fe/alicebalard-parasiteLoad-1b43216/DESCRIPTION
## * preparing 'parasiteLoad':
## * checking DESCRIPTION meta-information ... OK
## * checking for LF line-endings in source and make files and shell scripts
## * checking for empty or unneeded directories
## * building 'parasiteLoad_0.1.0.tar.gz'
```

Applying Alice's package on every gene

```
x <- field$IDO1



# Define function to be used to test, get the log lik and aic
```

```
tryDistrib <- function(x, distrib){
  # deals with fitdistr error:
  fit <- tryCatch(MASS::fitdistr(x, distrib), error=function(err) "fit failed")
  return(list(fit = fit,
              loglik = tryCatch(fit$loglik, error=function(err) "no loglik computed"),
              AIC = tryCatch(fit$aic, error=function(err) "no aic computed")))
}



findGoodDist <- function(x, distribs, distribs2){
  l =lapply(distribs, function(i) tryDistrib(x, i))
  names(l) <- distribs
  print(l)
  listDistr <- lapply(distribs2, function(i){
    if (i %in% "t"){
      fitdistrplus::fitdist(x, i, start = list(df =2))
    } else {
      fitdistrplus::fitdist(x,i)
    }}
  )
  par(mfrow=c(2,2))
  denscomp(listDistr, legendtext=distribs2)
  cdfcomp(listDistr, legendtext=distribs2)
  qqcomp(listDistr, legendtext=distribs2)
  ppcomp(listDistr, legendtext=distribs2)
  par(mfrow=c(1,1))
}
```

```
tryDistrib(x, "normal")
```

Functions for testing distributions

```
## $fit
##       mean            sd
##    15.1432029     4.3044978
##   ( 0.2278182) ( 0.1610918)
##
## $loglik
## [1] -1027.66
##
## $AIC
## NULL
```

```
tryDistrib(x, "binomial")
```

```
## $fit
## [1] "fit failed"
##
## $loglik
## [1] "no loglik computed"
##
## $AIC
## [1] "no aic computed"
```

```
tryDistrib(x, "student")
```

```
## $fit
```

```
## [1] "fit failed"
##
## $loglik
## [1] "no loglik computed"
##
## $AIC
## [1] "no aic computed"
```

```r
tryDistrib(x, "weibull")
```

```
## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## $fit
##      shape        scale
##    3.6718724   16.7658776
##  ( 0.1428414) ( 0.2566000)
##
## $loglik
## [1] -1032.752
##
## $AIC
## NULL
```

```r
tryDistrib(x, "weibullshifted")
```

```
## $fit
## [1] "fit failed"
##
## $loglik
## [1] "no loglik computed"
##
## $AIC
## [1] "no aic computed"
```

Write a function to apply alice's package on each gene to test for hybrid differences

```r
parasite_load_genes <- function(x) {
  field$Sex <- as.factor(field$Sex)

speparam <- parasiteLoad::getParamBounds("weibull", data = field, response = toString(x))


##All
print(parasiteLoad::analyse(data = field,
                        response = toString(x),
                        model = "weibull",
                        group = "Sex"))

}
```