

# RT-DC Data

Setup

```
PROJ_DIR = "/home/ericbarnhill/Documents/code/R/masscyto_R"
knitr::opts_chunk$set(root.dir = PROJ_DIR)
setwd(PROJ_DIR)
packages <- c("gdata", "tidyr", "dplyr", "magrittr", "ggplot2")
lapply(packages, require, character.only = T)
```

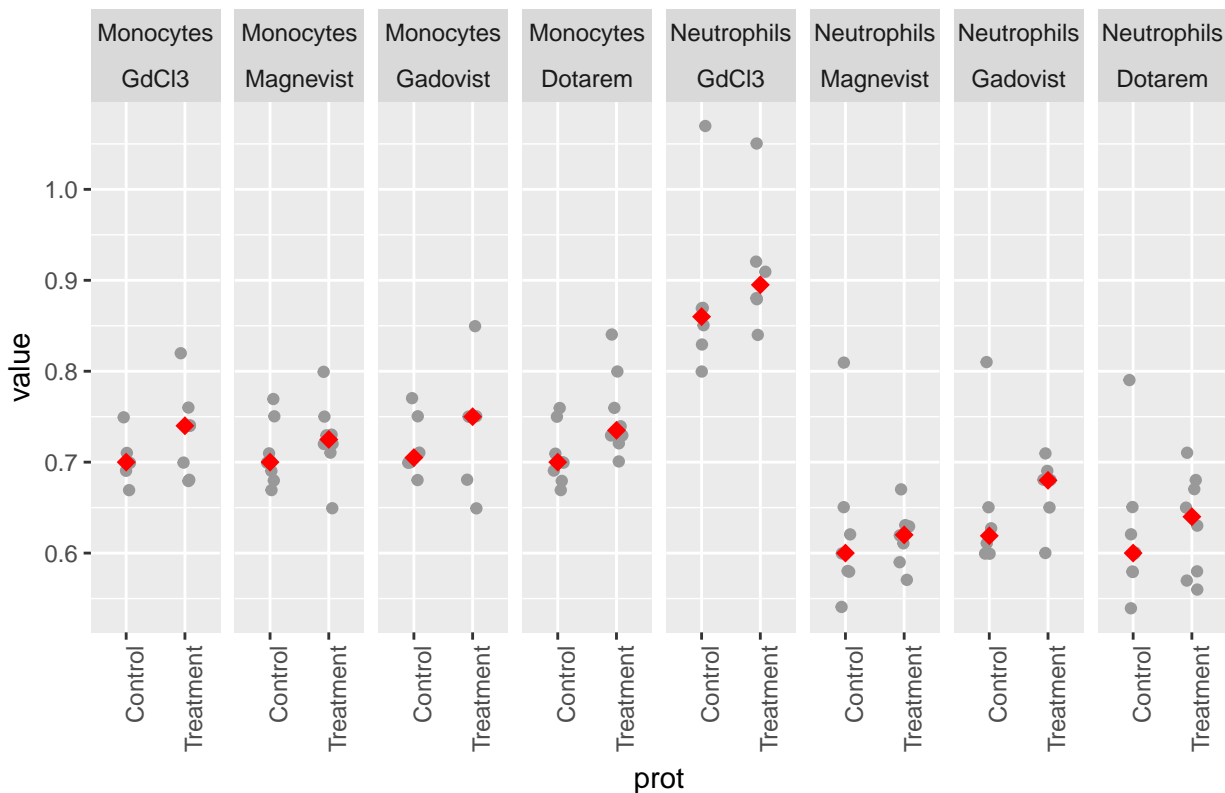
Load and clean data

```
source("load_rtdc.R")
XL_PATH <- file.path(PROJ_DIR, "rt-dc_eb.xls")
rtdc_tall <- load_rtdc(XL_PATH)
```

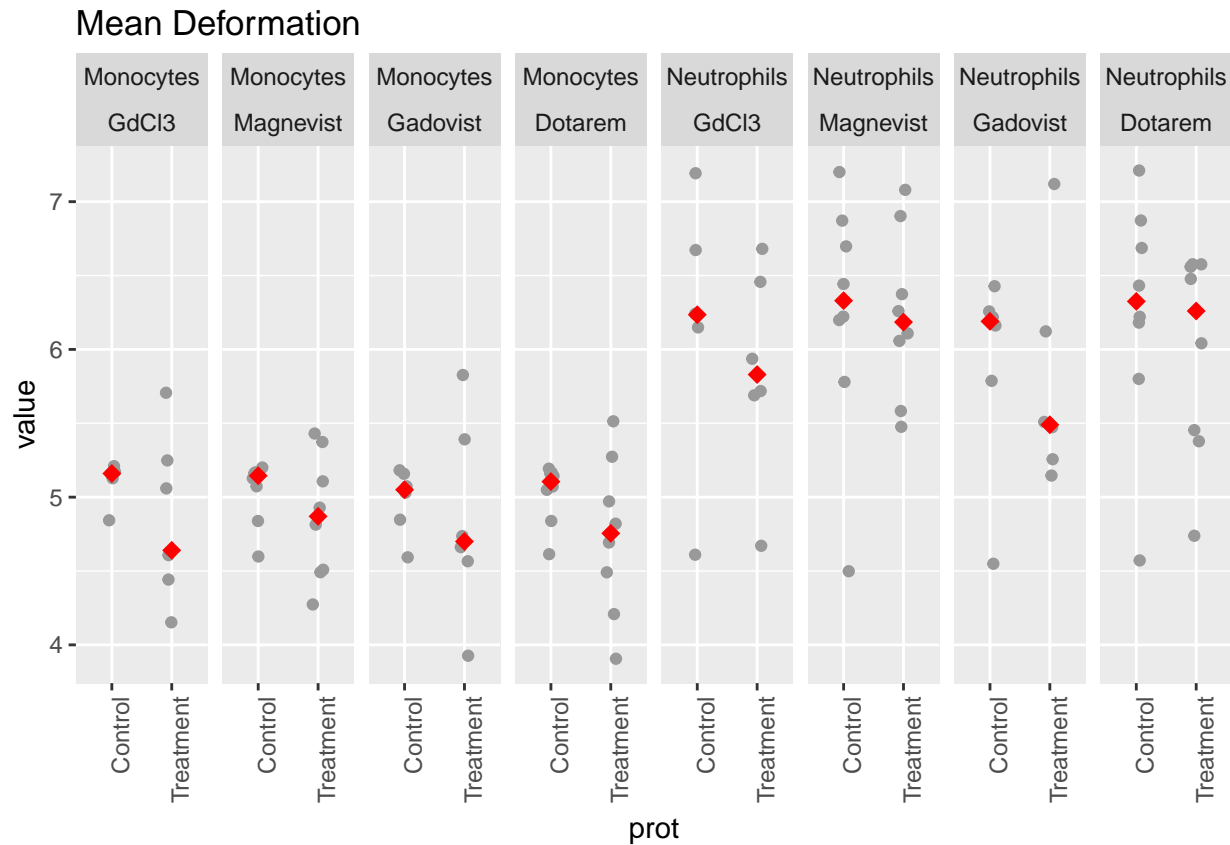
Explore summary stats of Young's modulus and deformation. There are many outliers so the effect is clearest using a robust statistic (median):

```
FUN = median
LOG = F;
plt_e <- rtdc_exploratory_boxplot(rtdc_tall, "mean_e", "Mean Young's Modulus", FUN, LOG)
print(plt_e)
```

## Mean Young's Modulus



```
plt_def <- rtdc_exploratory_boxplot(rtdc_tall, "mean_def", "Mean Deformation", FUN, LOG)
print(plt_def)
```

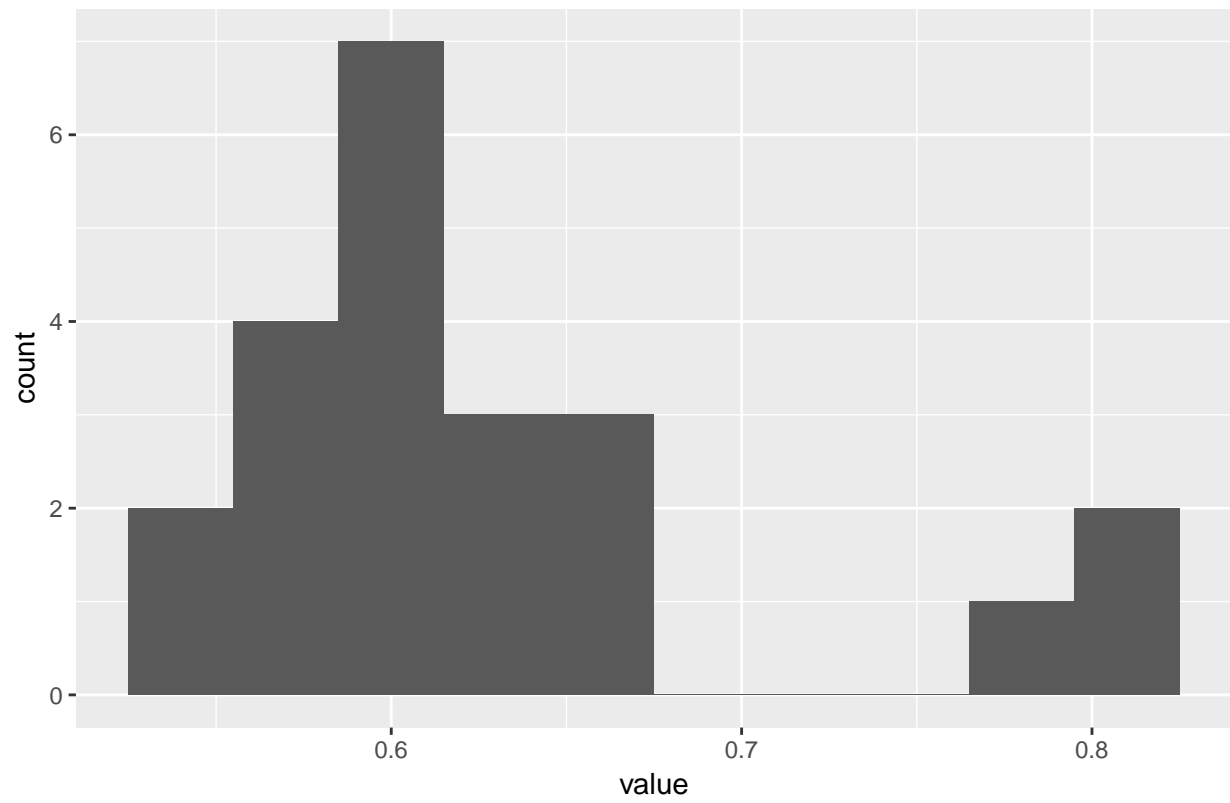


Log results look very similar, also the data is not probably not a good fit for the log: deformation has a range of (0,1) and Young's modulus can be negative.

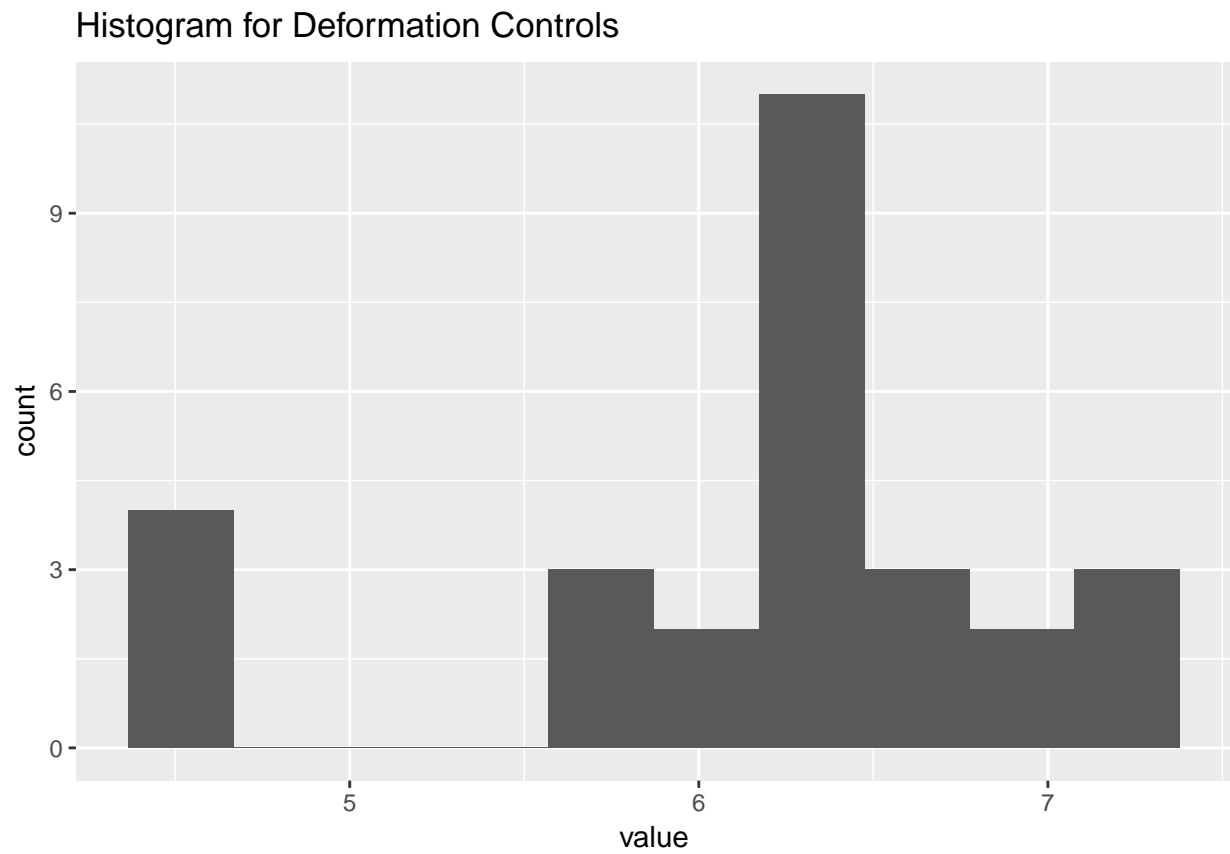
To get a sense of the distributions, look at histograms of controls (excluding GdCl3 Neutrophils Young's which seems to be off somehow) for one cell type

```
controls_e <- rtdc_tall %>%
  subset(prot == "Control" & measurement == "mean_e" & cell_type == "Neutrophils" & !(cell_type == "Neutrophils" & prot == "Control" & measurement == "mean_e"))
controls_def <- rtdc_tall %>%
  subset(prot == "Control" & measurement == "mean_def" & cell_type == "Neutrophils")
plt_e <- ggplot(controls_e) + geom_histogram(aes(x=value), bins=10) + ggtitle("Histogram for E Controls")
plt_def <- ggplot(controls_def) + geom_histogram(aes(x=value), bins=10) + ggtitle("Histogram for Deformation Controls")
print(plt_e)
```

Histogram for E Controls



```
print(plt_def)
```



Both E and deformation have outliers. Looking back at the strip plot, each measurement appears to have one heavy outlier. The E is reasonably normal. The deformation looks sparse and this needs to be understood.