First Tothill analysis

Ariel Hippen

2023-02-01

Contents

Having run the TCGA tothill data through BayesPrism, this notebook compares the samples' cell type composition with their subtype annotations from the Way pipeline and the patients' survival status/time.

```
suppressPackageStartupMessages({
    library(data.table)
    library(SingleCellExperiment)
    library(dplyr)
    library(yaml)
    library(stringr)
    library(ggplot2)
    library(survival)
    library(ggfortify)
})

params <- read_yaml("../../config.yml")
data_path <- params$data_path
local_data_path <- params$local_data_path
plot_path <- params$plot_path
```

```
tothill <- fread(paste(local_data_path, "deconvolution_output",
                       "tothill_default_bayesprism_results.tsv", sep = "/"))
tothill_melt <- melt(tothill)
```

```
## Warning in melt.data.table(tothill): id.vars and measure.vars are internally
## guessed when both are 'NULL'. All non-numeric/integer/logical type columns are
## considered id.vars, which in this case are columns [cell_type, ...]. Consider
## providing at least one of 'id' or 'measure' vars in future.
```

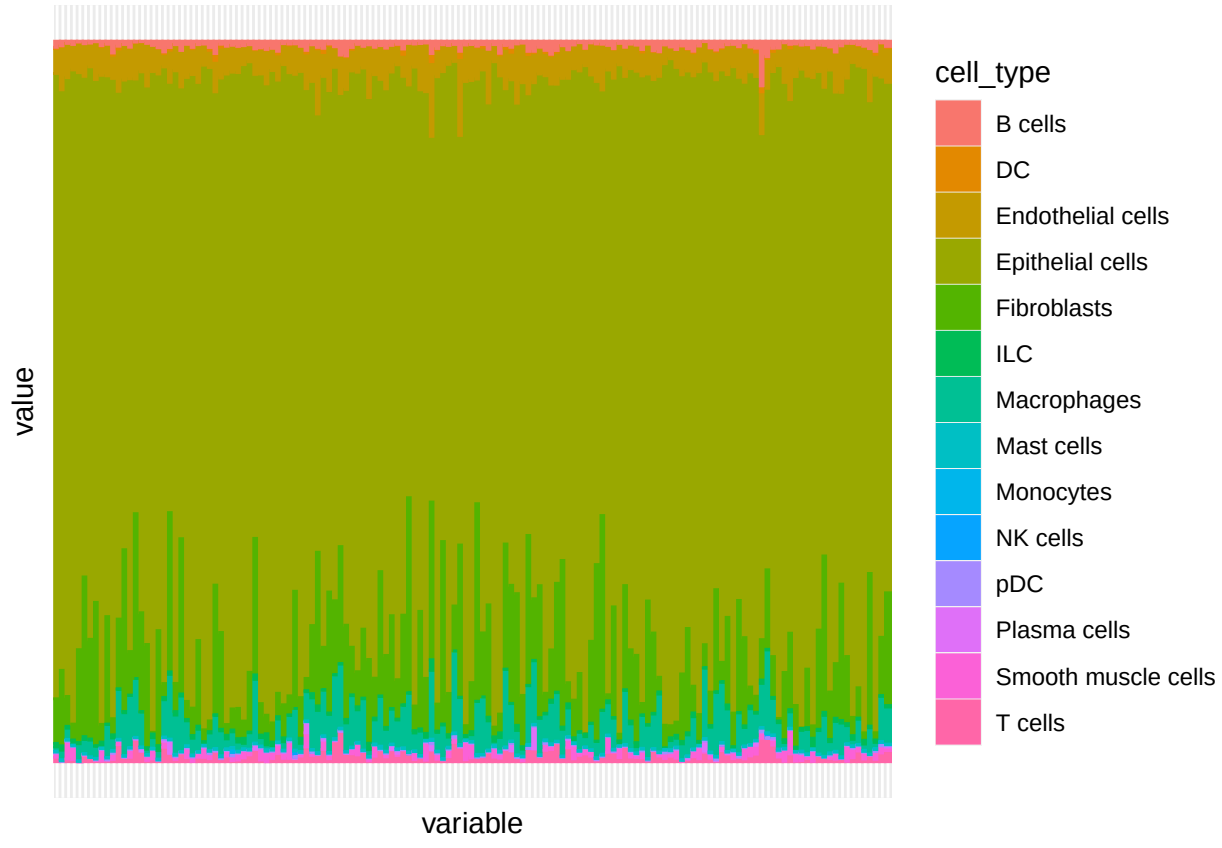
Cell composition

```
g <- ggplot(tothill_melt, mapping = aes(x=variable, y=value, fill=cell_type, color=cell_type)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x=element_blank(), #remove x axis labels
        axis.ticks.x=element_blank(), #remove x axis ticks
        axis.text.y=element_blank(),  #remove y axis labels
        axis.ticks.y=element_blank()) #remove y axis ticks

plotfile <- paste(plot_path, "evaluation_plots",
                  "tothill_proportion_barchart.png", sep = "/")
png(filename = plotfile, width = 1200)
g
dev.off()
```

```
## pdf
##   2
```

```
g
```




```
# Switch so cell types are columns and samples are rows for easier analysis
cell_types <- tothill$cell_type

tothill$cell_type <- NULL
tothill_t <- t(as.matrix(tothill))
colnames(tothill_t) <- cell_types
tothill_t <- as.data.frame(tothill_t)
```

Cell composition by survival

```
# Load survival data
tothill_survival <- fread(paste(local_data_path, "Tothill", "tothill_survival.tsv",
                                sep = "/"))
```

```
# Combine survival data with %
tothill_t$unique_patient_ID <- rownames(tothill_t)
tothill_master <- full_join(tothill_survival, tothill_t)
```

```
## Joining, by = "unique_patient_ID"
```

```
tothill_master$Immune <- tothill_master$Macrophages + tothill_master$Monocytes + tothill_master$`Plasma
  tothill_master$DC + tothill_master$`NK cells` + tothill_master$pDC + tothill_master$`B cells` + tothil
  tothill_master$`Mast cells`
```

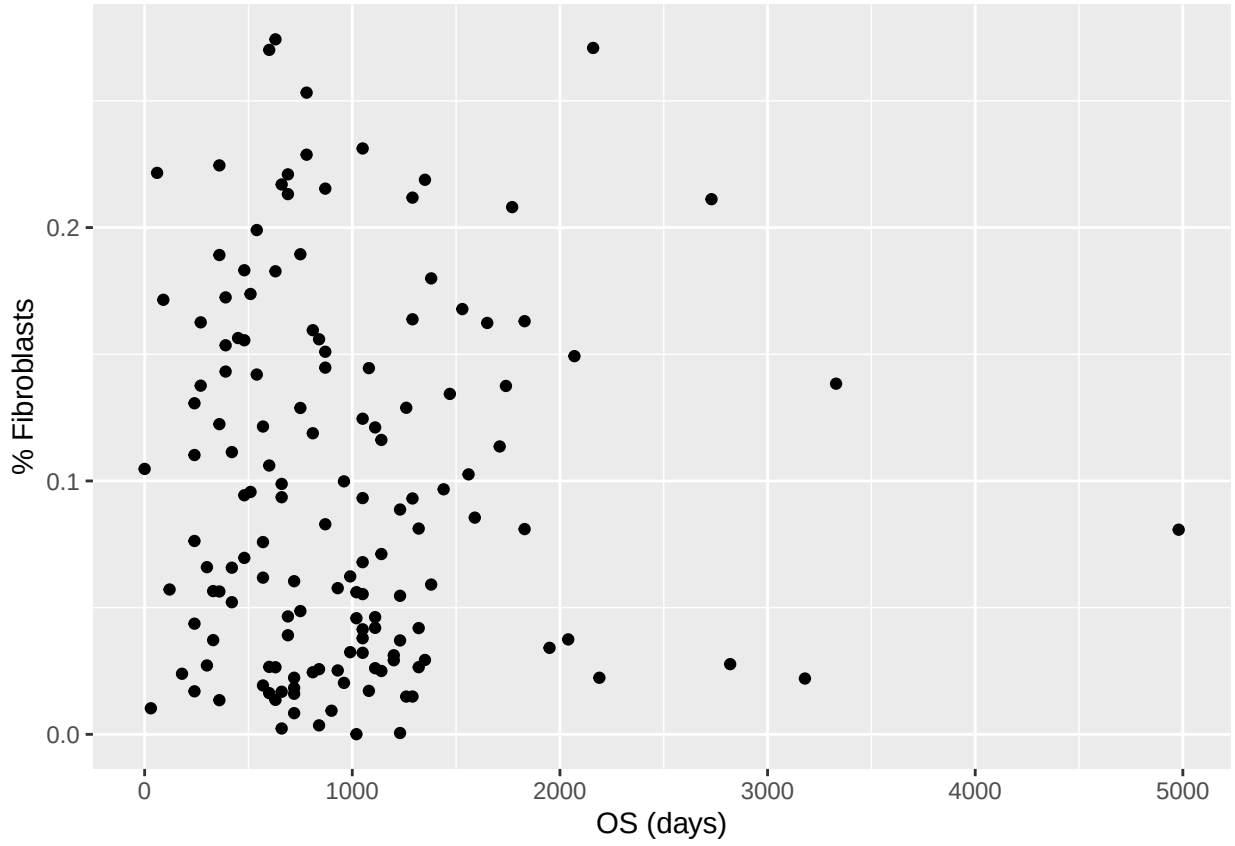
```
g <- ggplot(tothill_master, mapping = aes(x=tothill_master$days_to_death, y=tothill_master$Fibroblasts))
  geom_point() + xlab("OS (days)") + ylab("% Fibroblasts")
plotfile <- paste(plot_path, "evaluation_plots", "tothill_survival_by_fibroblasts.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## pdf
##   2
```

```
g
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



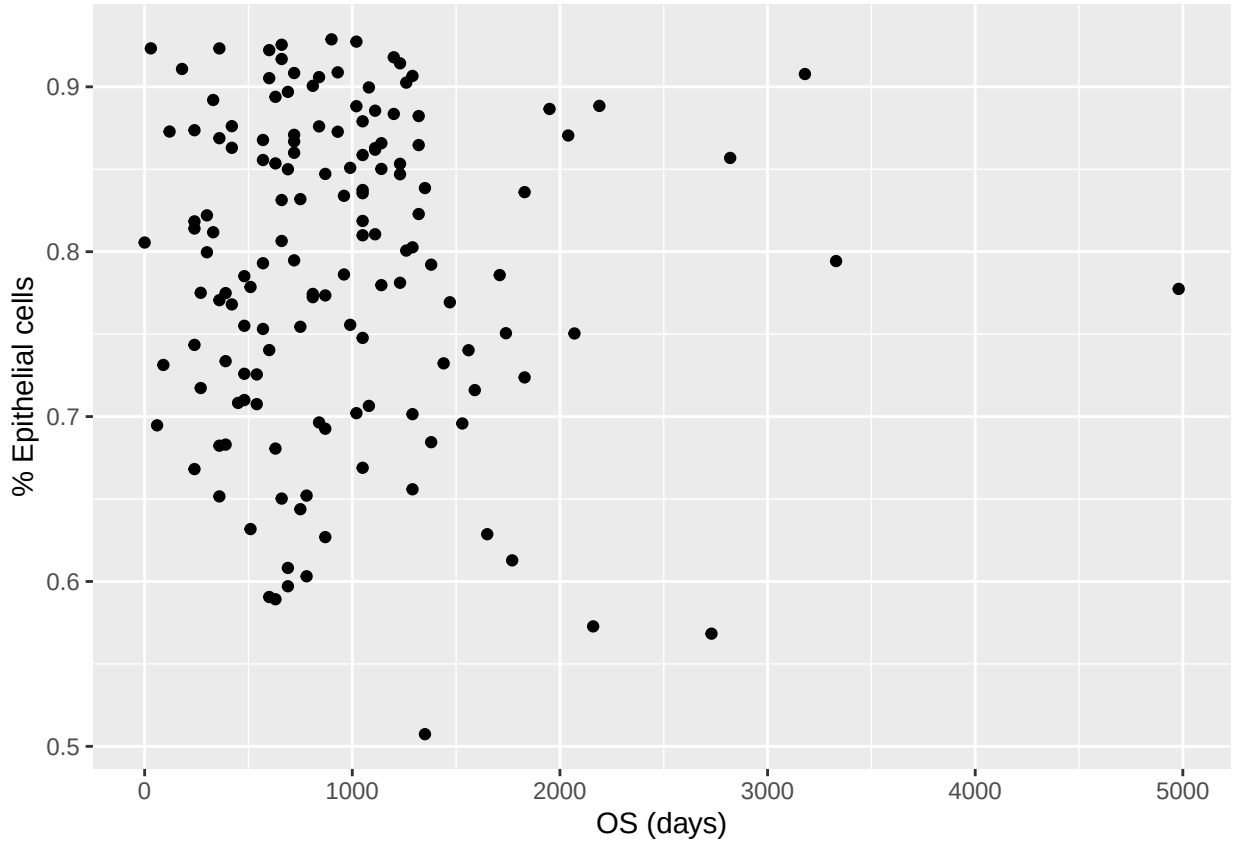

```
g <- ggplot(tothill_master, mapping = aes(x=tothill_master$days_to_death, y=tothill_master$`Epithelial
  geom_point() + xlab("OS (days)") + ylab("% Epithelial cells")
plotfile <- paste(plot_path, "evaluation_plots", "tothill_survival_by_epithelial.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## pdf
##   2
```

```
g
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



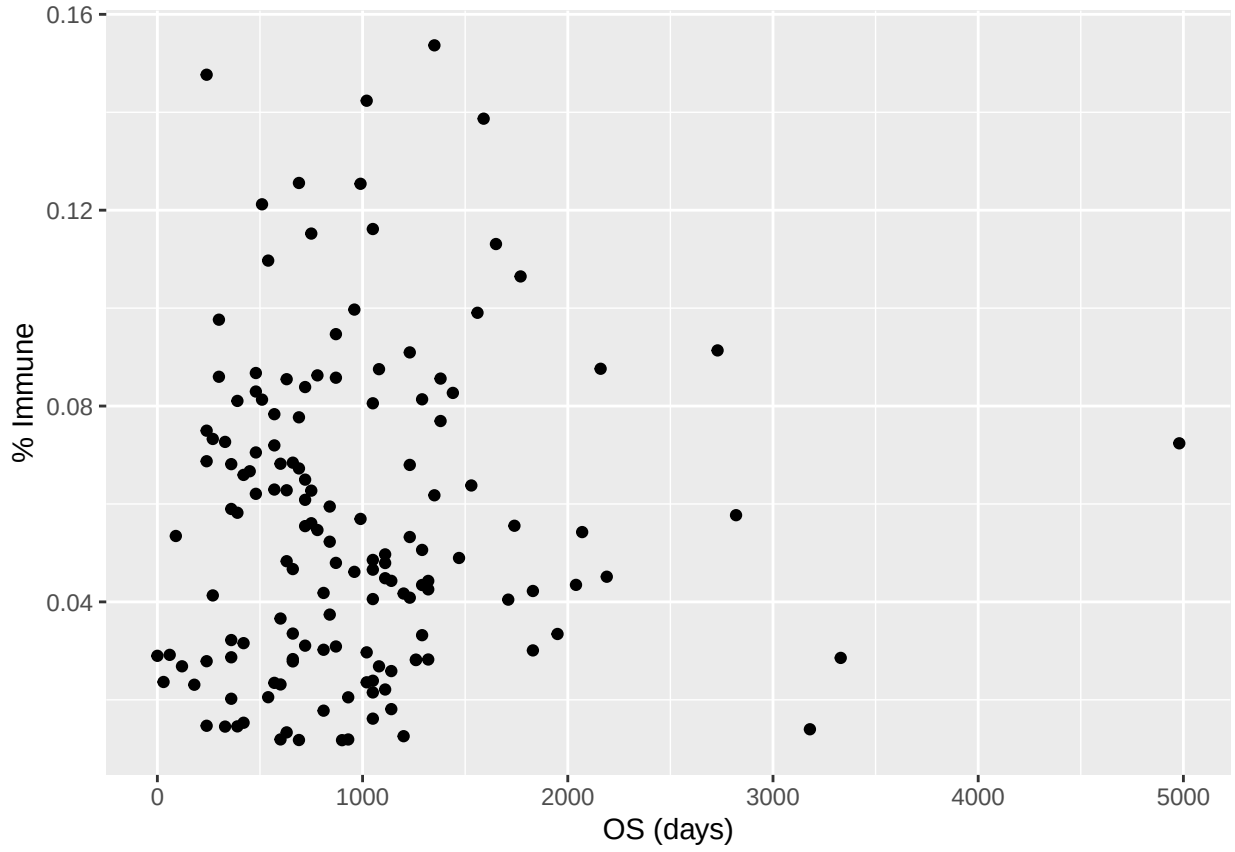

```
g <- ggplot(tothill_master, mapping = aes(x=tothill_master$days_to_death, y=tothill_master$Immune)) +
  geom_point() + xlab("OS (days)") + ylab("% Immune")
plotfile <- paste(plot_path, "evaluation_plots", "tothill_survival_by_immune.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## pdf
##   2
```

```
g
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



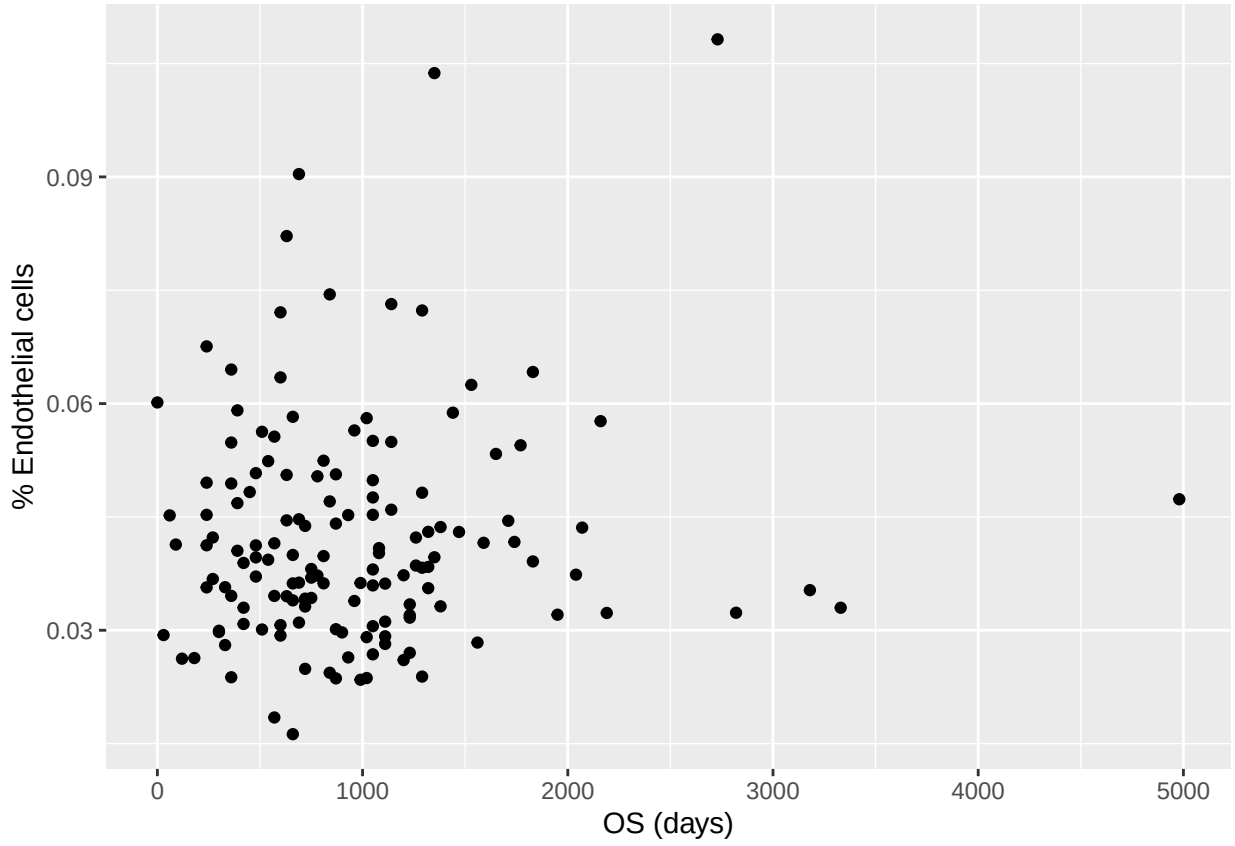

```
g <- ggplot(tothill_master, mapping = aes(x=tothill_master$days_to_death, y=tothill_master$`Endothelial
  geom_point() + xlab("OS (days)") + ylab("% Endothelial cells")
plotfile <- paste(plot_path, "evaluation_plots", "tothill_survival_by_endothelial_cells.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## pdf
##   2
```

```
g
```

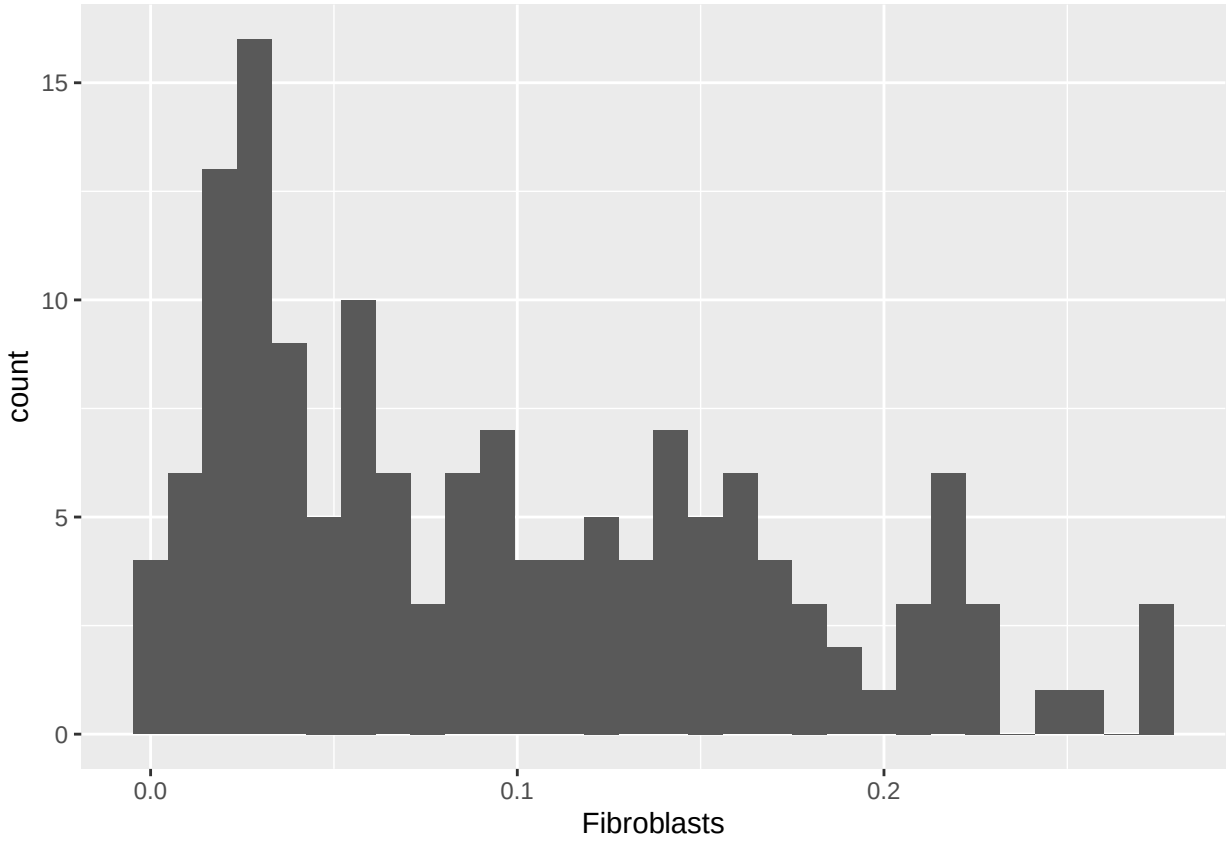
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



Kaplan Meier curves

```
# Put the samples into quartiles based on fibroblast content
ggplot(tothill_master, mapping = aes(x=Fibroblasts)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
quantiles <- quantile(tothill_master$Fibroblasts)
q1 <- quantiles[2]
q3 <- quantiles[4]
tothill_master$high_fibro <- ifelse(tothill_master$Fibroblasts > q3, 1, 0)
```

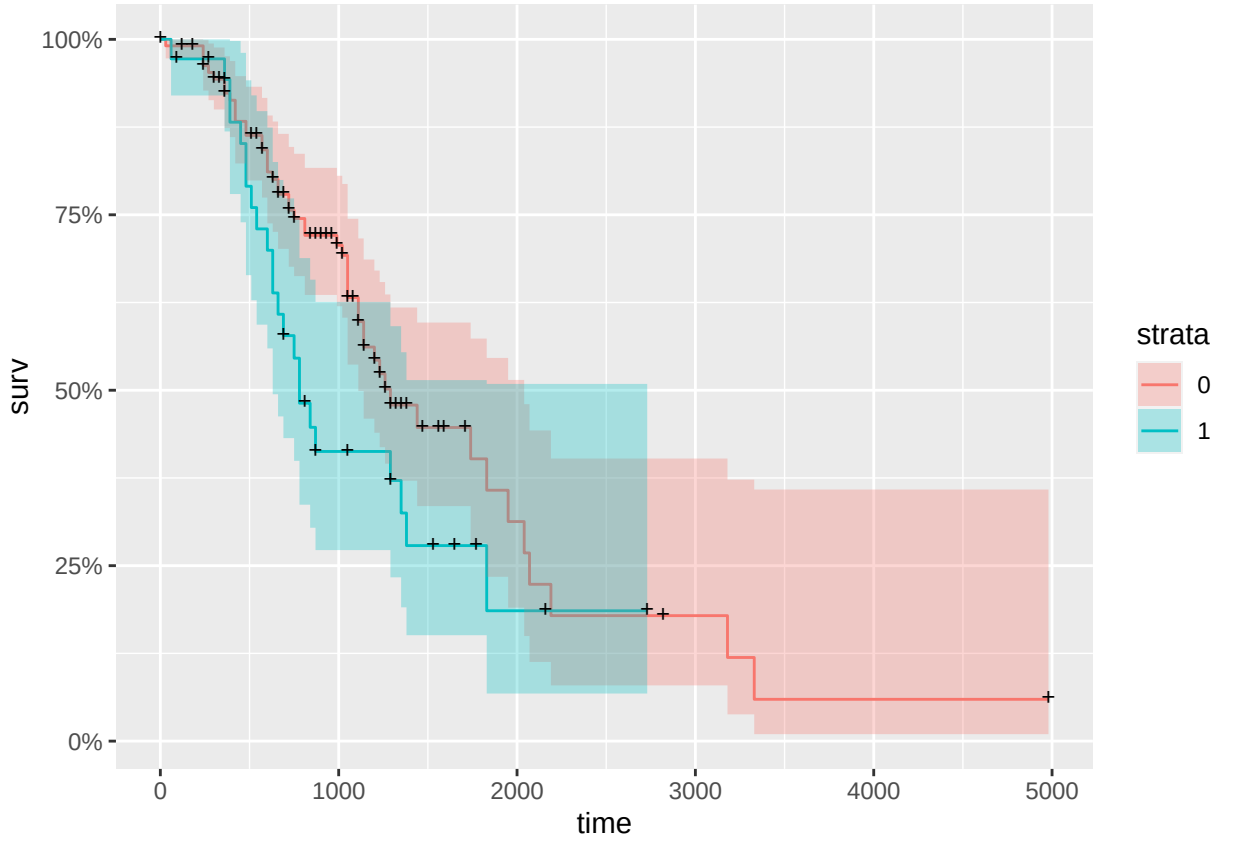
```
# Get Kaplan-Meier curves
tothill_master$vital_status <- recode(tothill_master$vital_status,
                                      "deceased"=1,
                                      "living"=0)

km <- Surv(time = tothill_master$days_to_death,  event = tothill_master$vital_status)
km_treatment<-survfit(km~high_fibro,data=tothill_master,type='kaplan-meier',conf.type='log')

plotfile <- paste(plot_path, "evaluation_plots", "tothill_KaplanMeier_fibroblasts.png", sep = "/")
png(filename = plotfile)
autoplot(km_treatment)
dev.off()
```

```
## pdf
##   2
```

```
autoplot(km_treatment)
```



Subtypes

```
# Get subtype annotations
cluster_file <- paste(local_data_path, "cluster_assignments", "FullClusterMembership.csv", sep = "/")
cluster_list <- fread(cluster_file)

setnames(cluster_list, "V1", "unique_patient_ID")

tothill_t <- left_join(tothill_t, cluster_list)
```

```
## Joining, by = "unique_patient_ID"
```

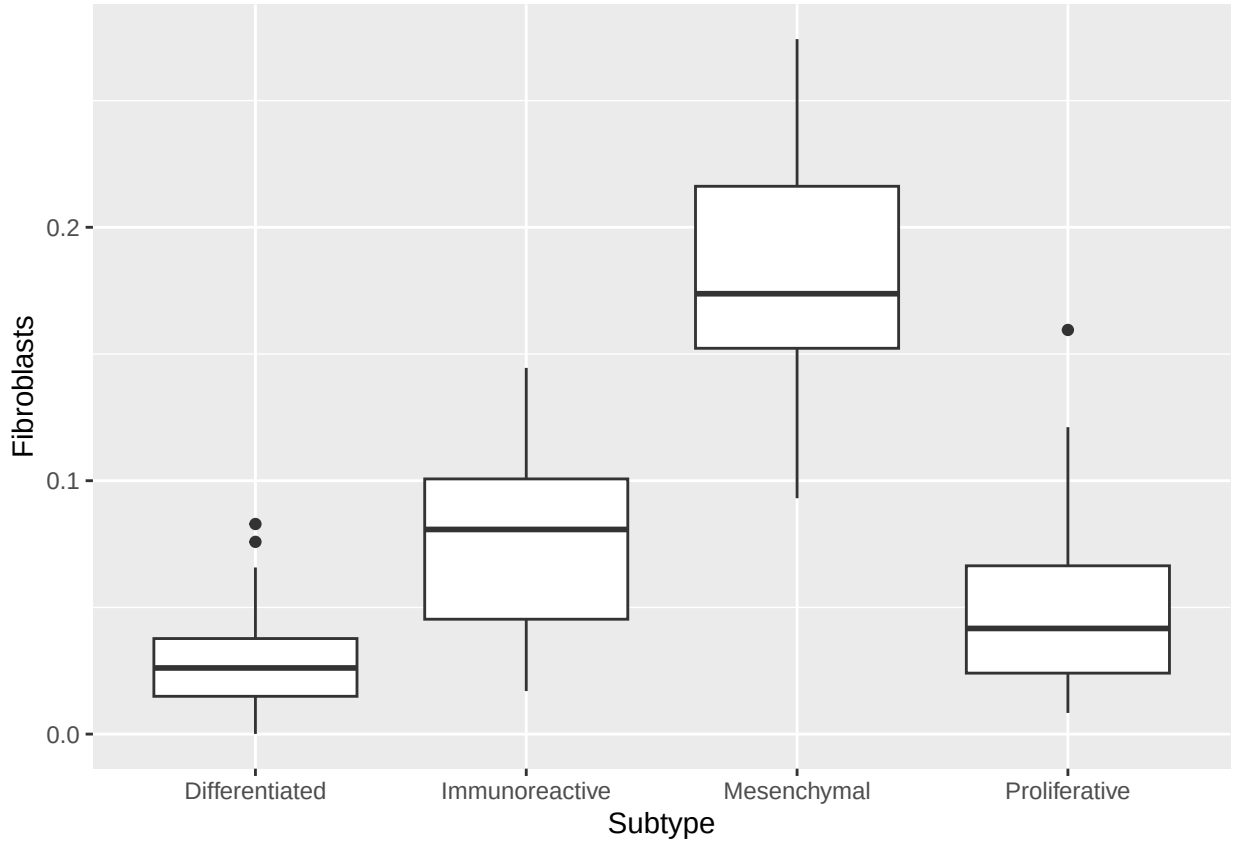
```
tothill_t$Subtype <- recode(tothill_t$ClusterK4_kmeans,
                            "1" = "Mesenchymal",
                            "2" = "Proliferative",
                            "3" = "Immunoreactive",
                            "4" = "Differentiated")

tothill_t$Immune <- tothill_t$`T cells` + tothill_t$Macrophages + tothill_t$Monocytes + tothill_t$`Plasm
  tothill_t$DC + tothill_t$`NK cells` + tothill_t$pDC + tothill_t$`B cells` + tothill_t$ILC + tothill_t$
```

```
# Compare cell type proportions of subtypes
g <- ggplot(tothill_t, mapping = aes(x=Subtype, y=Fibroblasts)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_fibroblasts_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

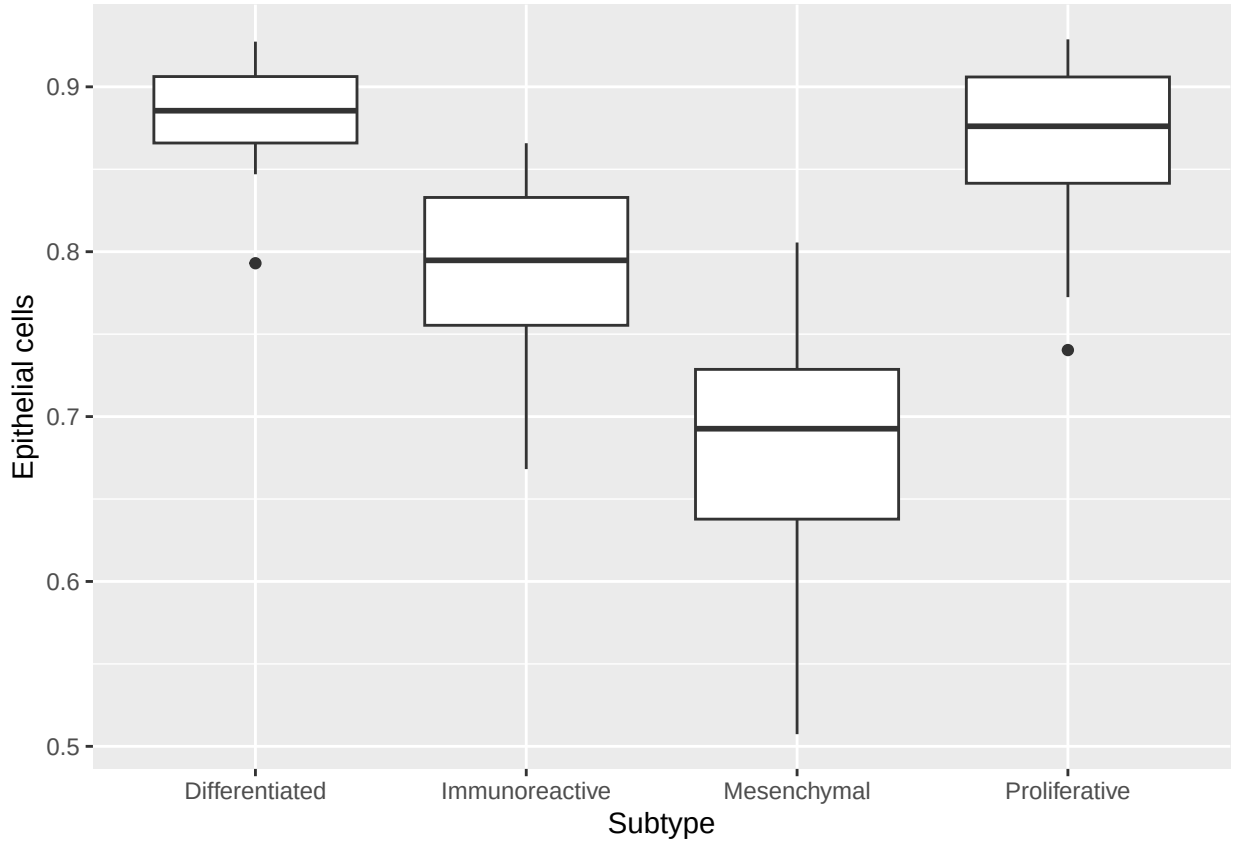
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=Subtype, y=`Epithelial cells`)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_epithelial_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

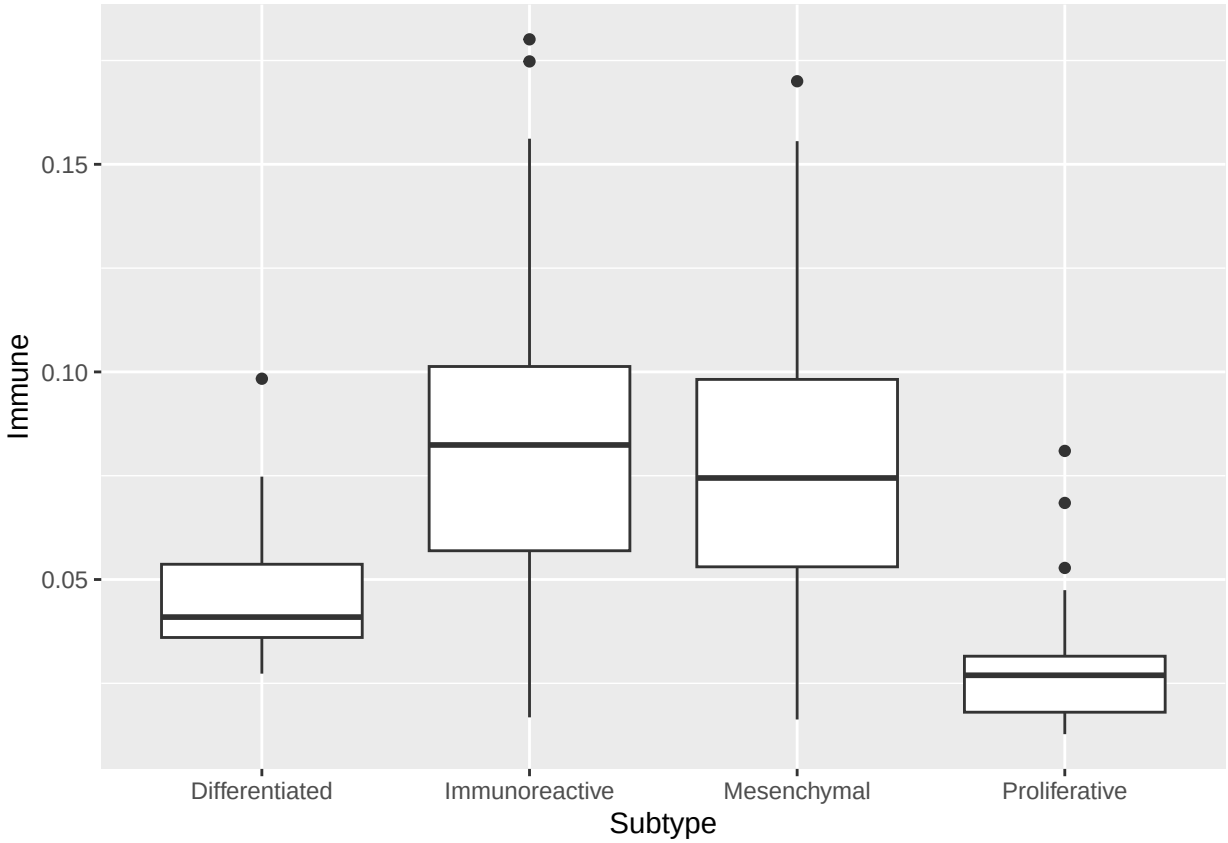
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=Subtype, y=Immune)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_immune_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

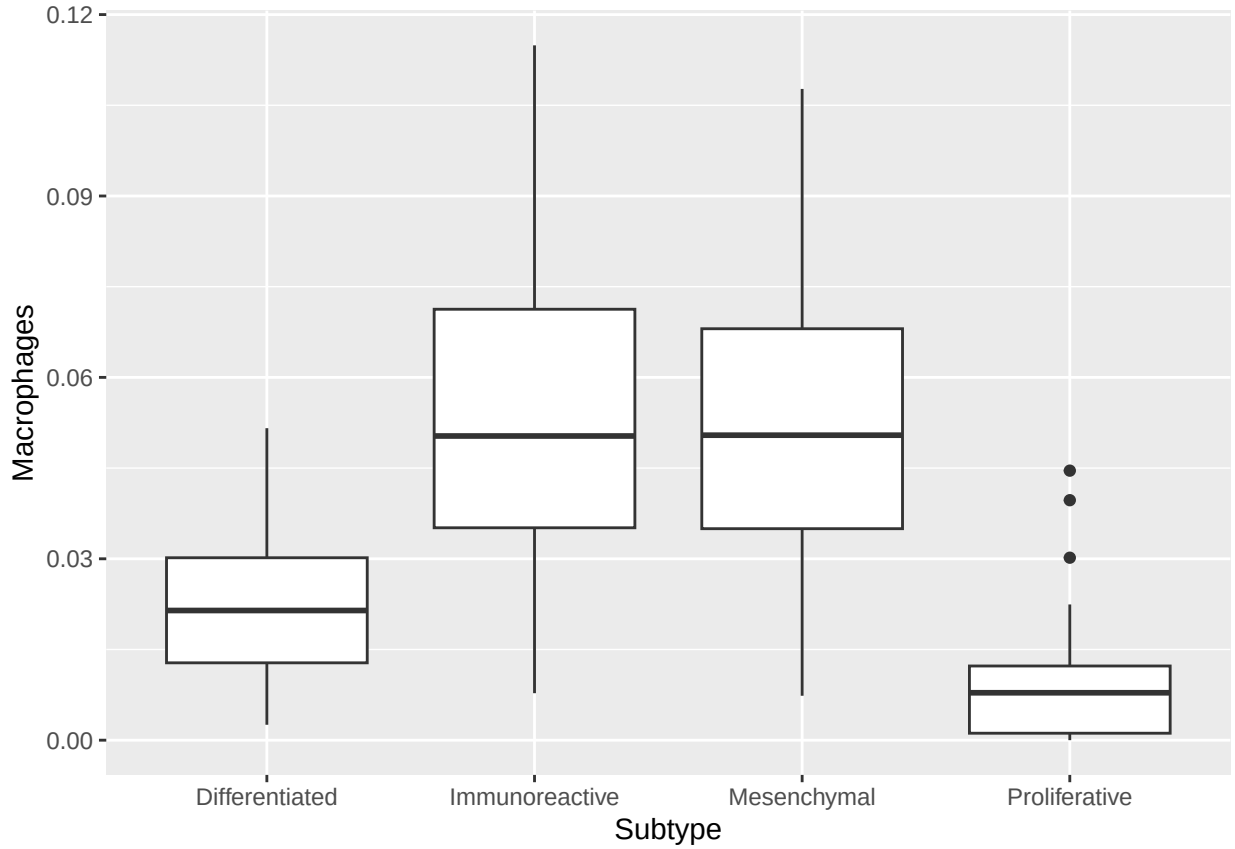
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=Subtype, y=Macrophages)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_macrophages_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

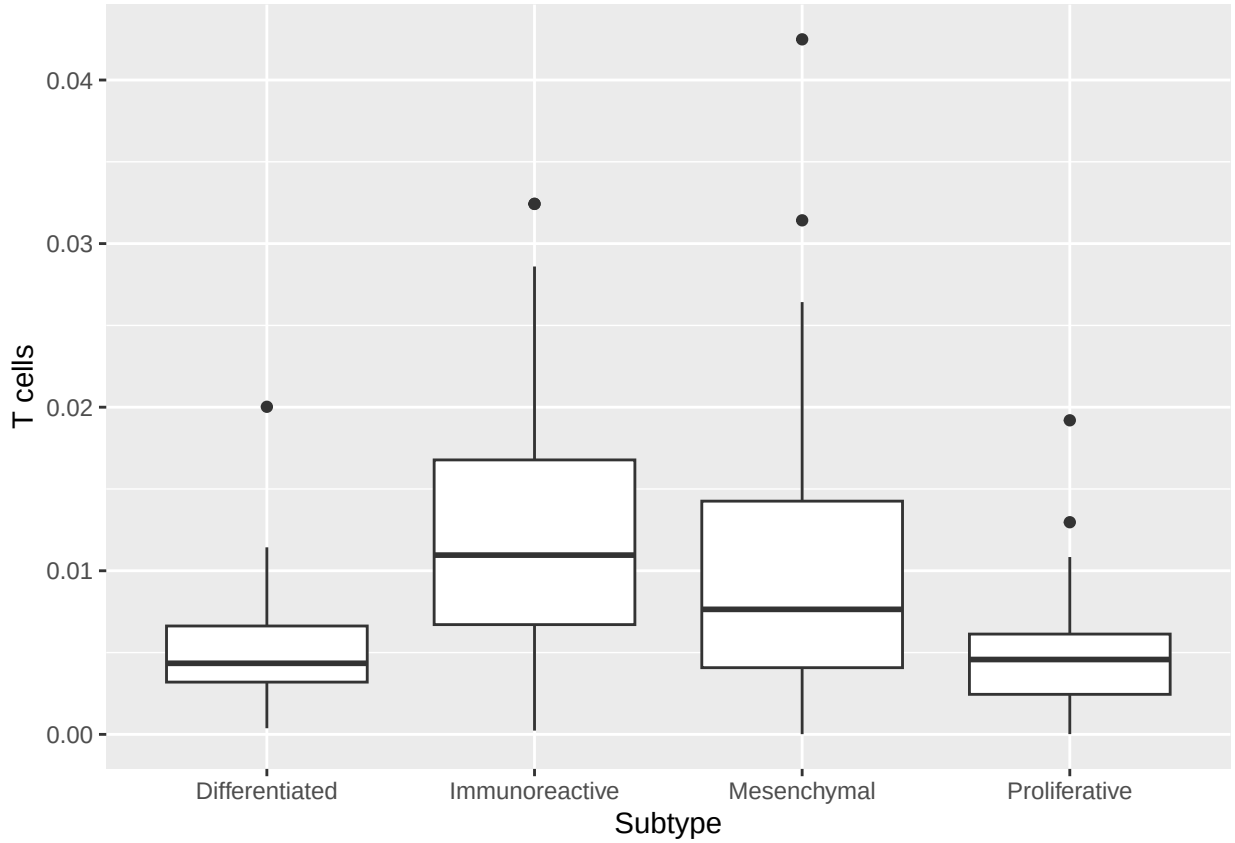
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=Subtype, y=`T cells`)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_tcells_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

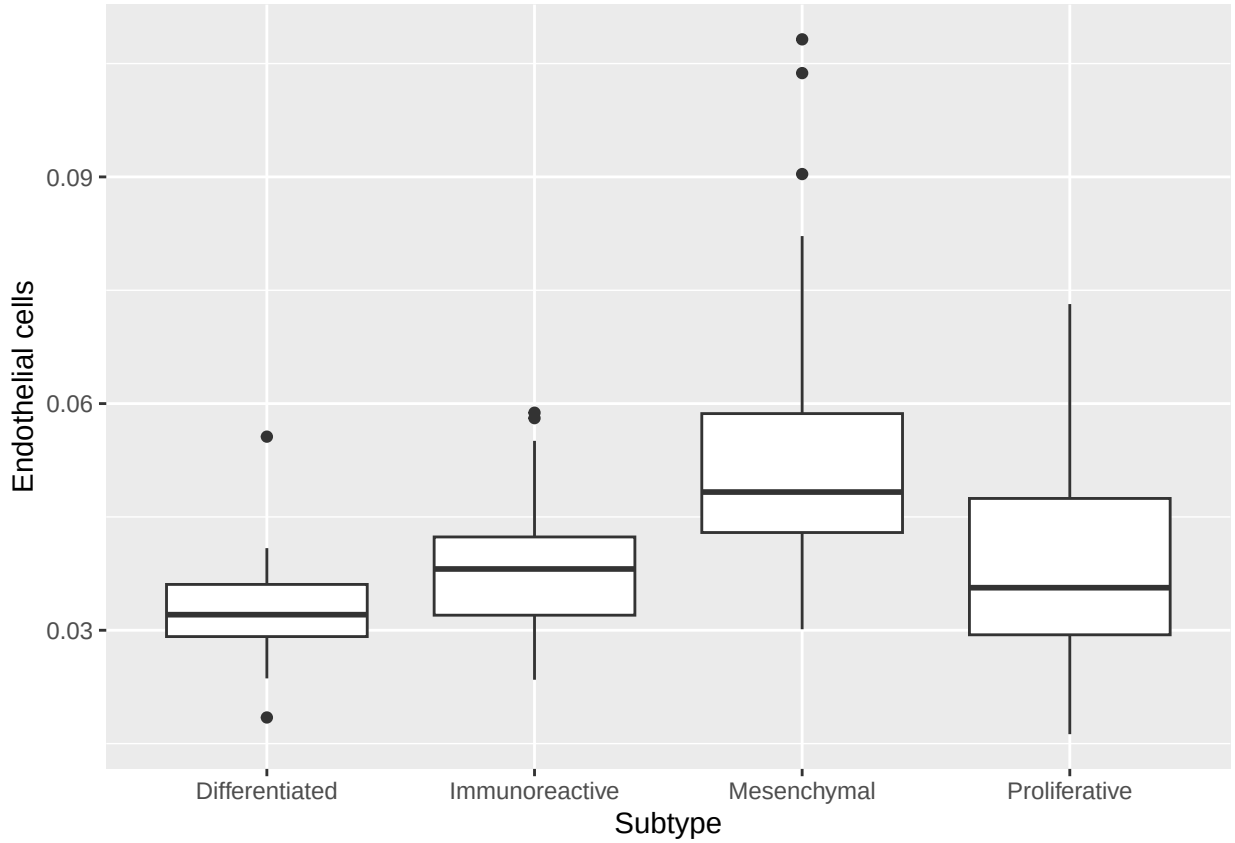
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=Subtype, y=`Endothelial cells`)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_endothelial_by_subtype.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

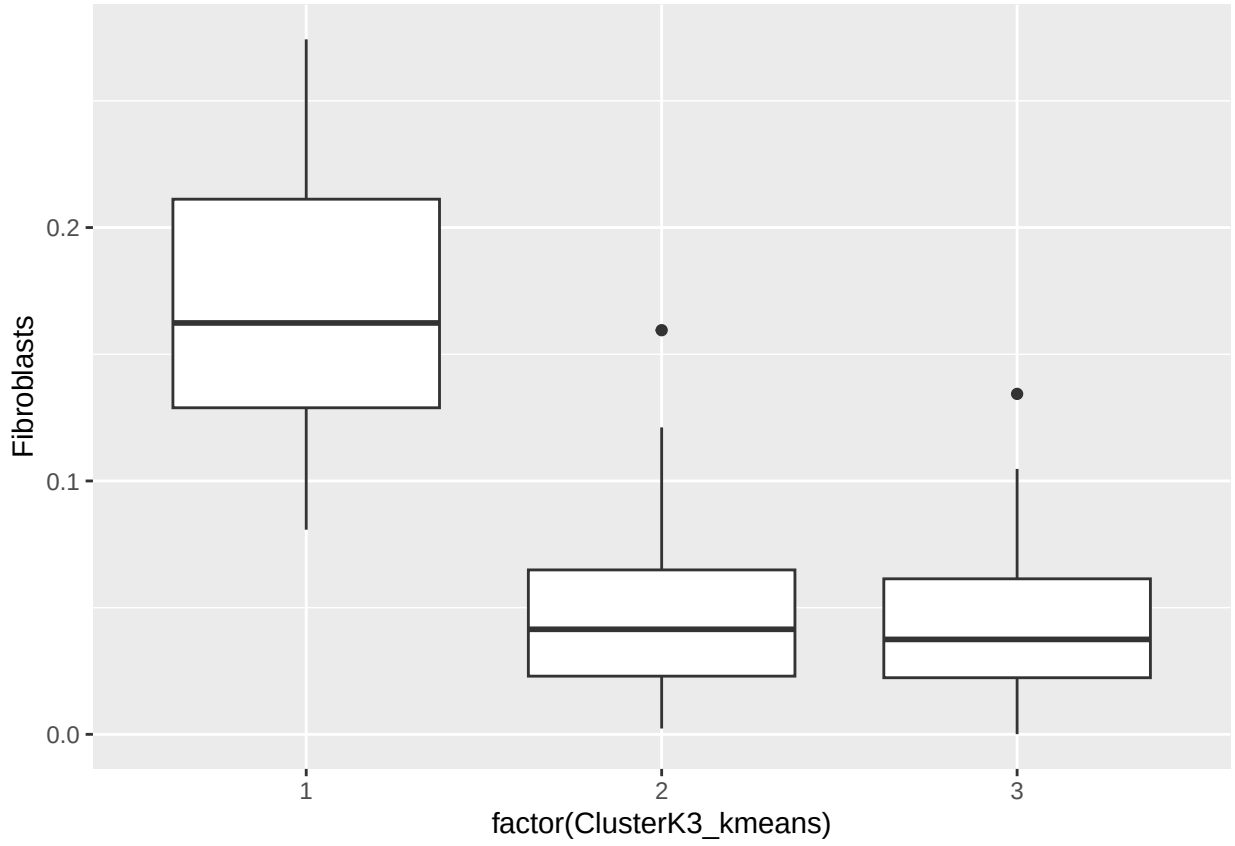
```
g
```




```
# Compare cell type proportions of subtypes for k=3
g <- ggplot(tothill_t, mapping = aes(x=factor(ClusterK3_kmeans), y=Fibroblasts)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_fibroblasts_by_subtype_k3.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

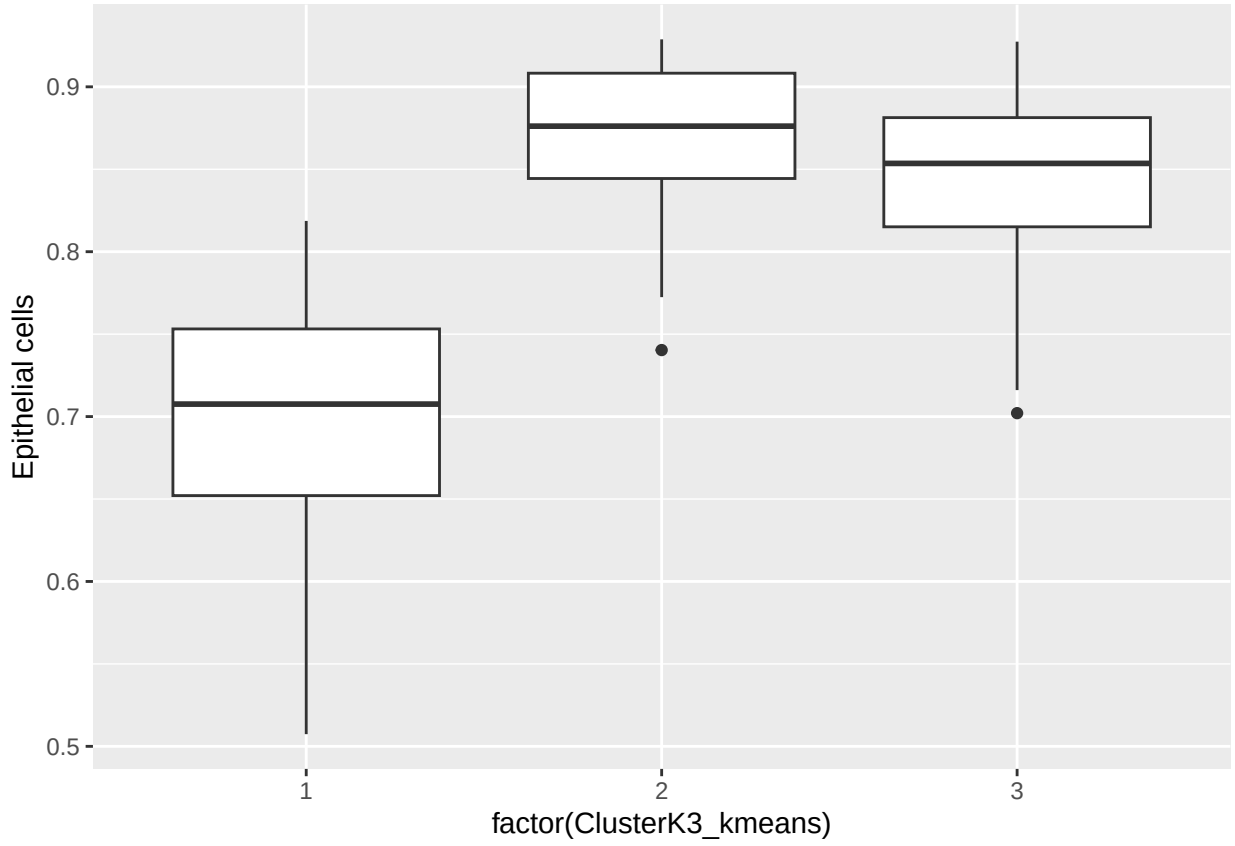
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=factor(ClusterK3_kmeans), y=`Epithelial cells`)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_epithelial_by_subtype_k3.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

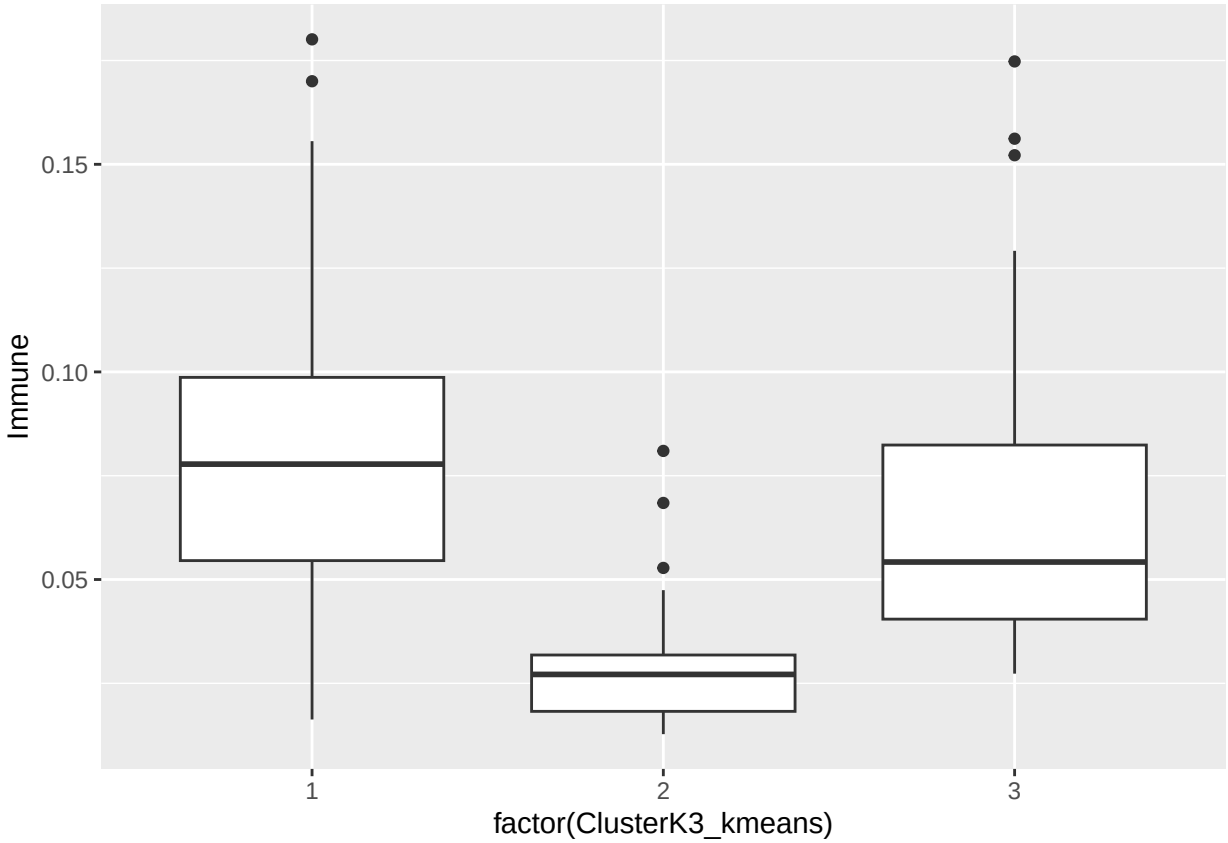
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=factor(ClusterK3_kmeans), y=Immune)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_immune_by_subtype_k3.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

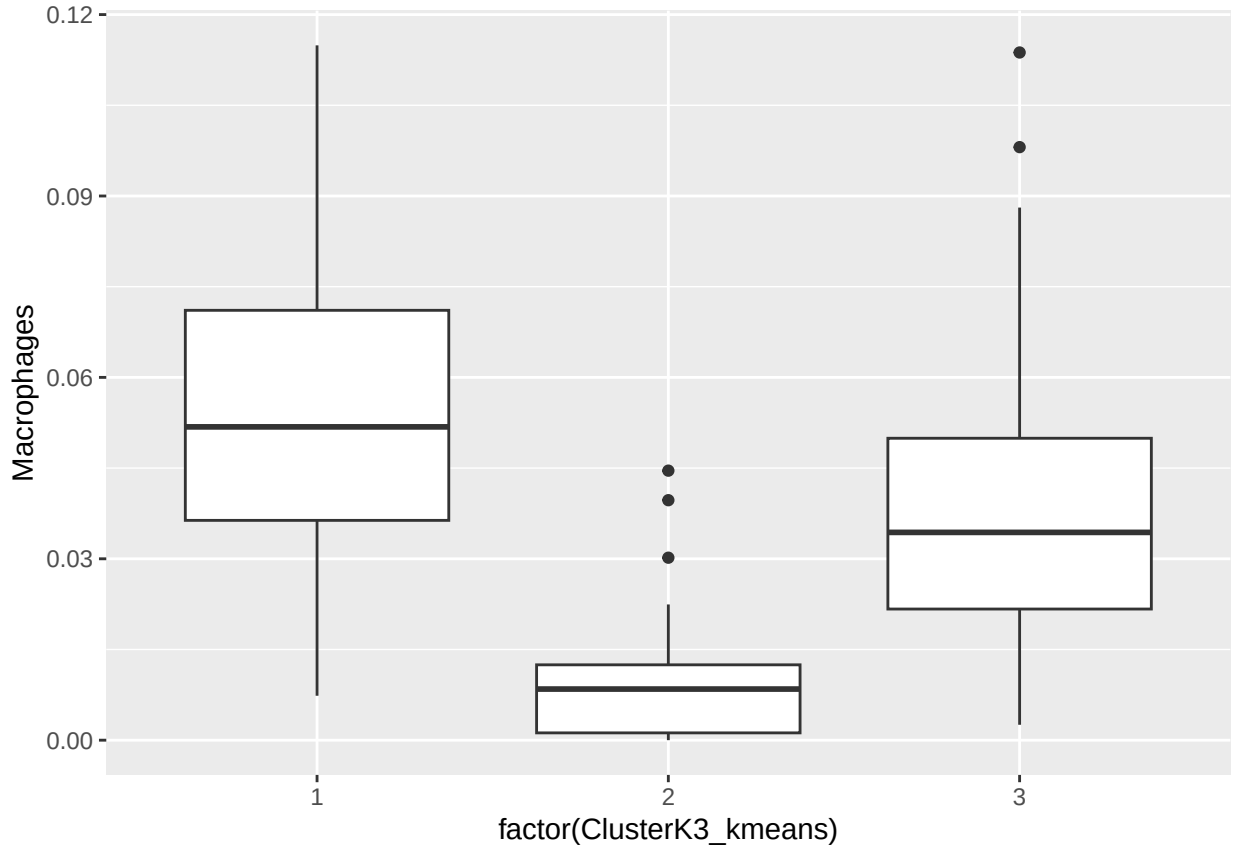
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=factor(ClusterK3_kmeans), y=Macrophages)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_macrophages_by_subtype_k3.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

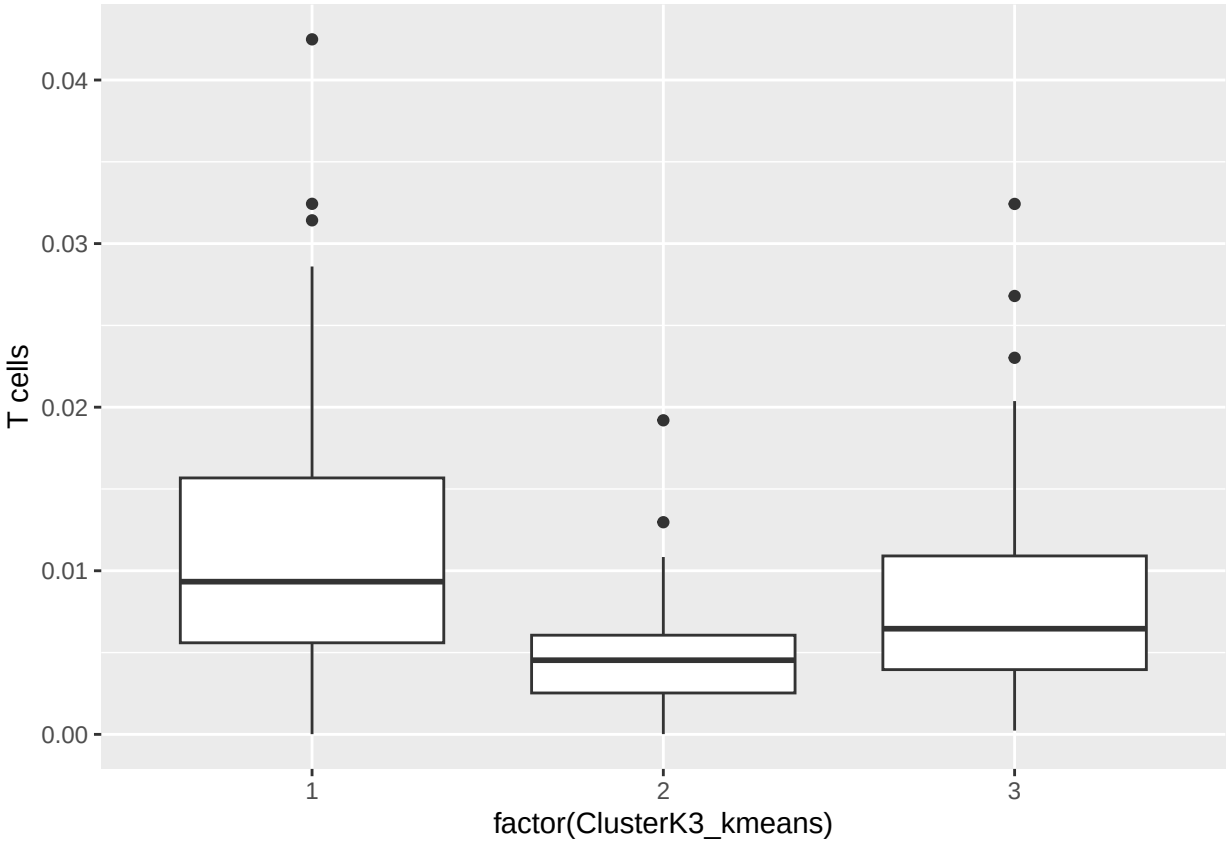
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=factor(ClusterK3_kmeans), y=`T cells`)) + geom_boxplot()
plotfile <- paste(plot_path, "evaluation_plots", "tothill_tcells_by_subtype_k3.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

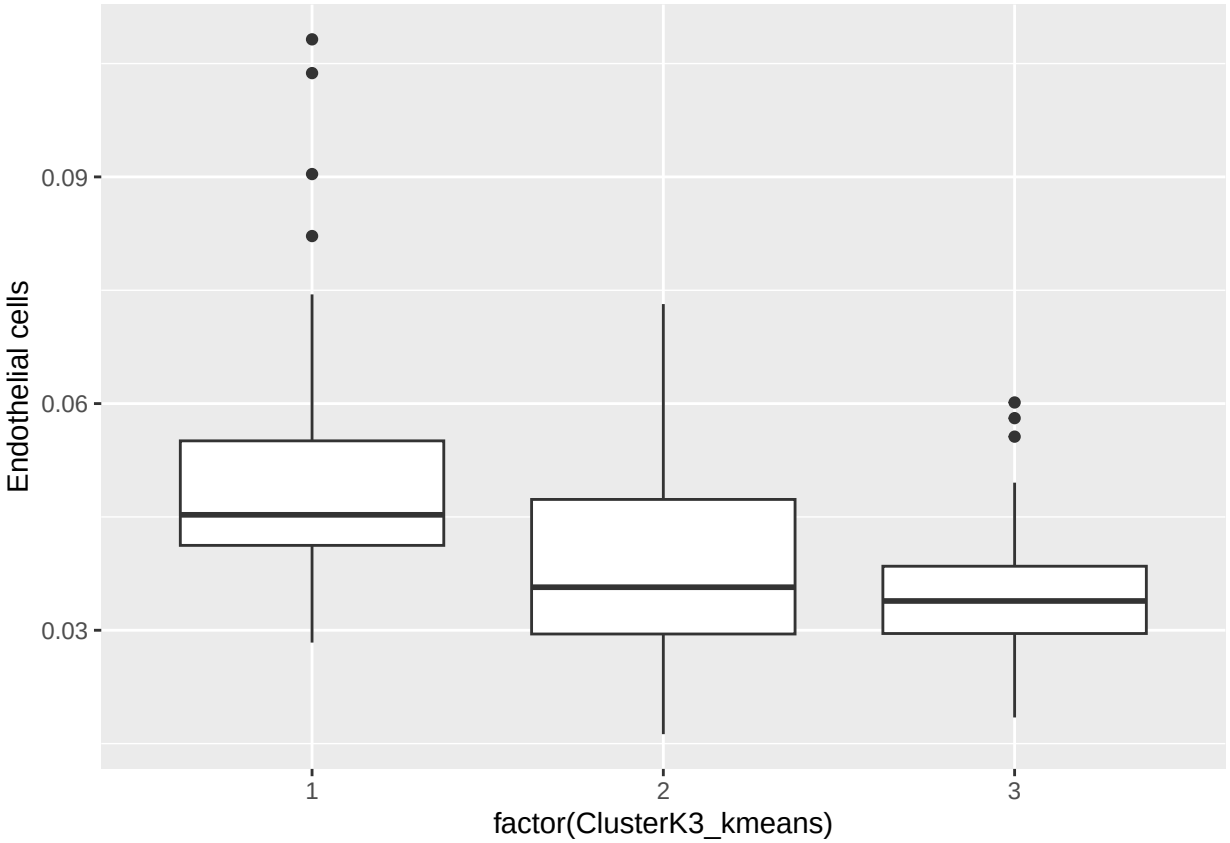
```
g
```




```
g <- ggplot(tothill_t, mapping = aes(x=factor(ClusterK3_kmeans), y=`Endothelial cells`)) + geom_boxplot
plotfile <- paste(plot_path, "evaluation_plots", "tothill_endothelial_by_subtype_k3.png", sep = "/")
png(filename = plotfile); g; dev.off()
```

```
## pdf
##   2
```

```
g
```


Endothelial cells
0.09
0.06
0.03
1
2
3
factor(ClusterK3_kmeans)