Identity within plots for soil variables

Here we present analysis of the "insitu" sheet of "soil_variables_DPE" table. The table represents data for the "Soil pH at 10cm depth", "Soil temperature, °C", "Active layer thickness, cm" and "Air temperature, °C" variables. Each variable representad by 5 observations within each of 3 blocks within each of 10 blocks.

We want to test if there are any significant differences within plots of each of the block. Besides, it is interesting, is there any significant differences in these variables between ridge and lakebed locations.

```
library(xlsx)
```

```
## Loading required package: rJava
```

```
## Loading required package: xlsxjars
```

```
library(ggplot2)
library(ggpubr)
```

```
## Loading required package: magrittr
```

```
library(plyr)
source("./functions.R")
```

```
ins_path <- "/home/lena/Tundra_exp/soil_variables_DPE.xlsx"
ins <- read.xlsx(ins_path, 1)  # read first sheet

clmn = colnames(ins) %in% c("Blk", "Plot.", "Air.temperature...C.", "Soil_pH_10cm_depth", "Soil_temp_deg
ins = ins[!is.na(ins$Blk),clmn]
colnames(ins) = c("block", "plot", "air_temp", "soil_pH", "soil_temp", "ALT_cm")
ynames = colnames(ins)[3:ncol(ins)]
varnames = c("Air temperature, °C", "Soil pH at 10cm depth", "Soil temperature, °C", "Active layer thick
n_samp = 5              # number of observations in each plot
rm(clmn, ins_path)
```

Statistical tests

In order to test differences within plots, we perform ANOVA for plots that meet ANOVA's assumptions and Kruskal-Wallis for others

```
test = sapply(c(4:ncol(ins)), function(cl) {
  # ANOVA test
  anpv = sapply(c(0:9), function(x) {
    inexes_within_block = (1 + n_samp*3*x):(n_samp*3*x+n_samp*3)
    plant.mod1 = lm(paste0(colnames(ins)[cl]," ~ plot"), data = ins[ inexes_within_block ,])
    res = anova(plant.mod1)$`Pr(>F)`                                # results of ANOVA
    res[2] = shapiro.test(residuals(plant.mod1))$p.value            # normality of residuals
    res[3] = bartlett.test(ins[ inexes_within_block ,cl],
                      ins$plot[ inexes_within_block ])$p.value      # homogeneity of variances
    return(res)
  })
  non_anova_blocks = which((anpv[2, ] < 0.05) | (anpv[3, ] < 0.05))    # block's numbers that do not me
                                                                    # ANOVA's assumptions
  # Kruskal-Wallis for non-normal residuals
  krpv = sapply( non_anova_blocks , function(x) kruskal.test(ins[(1 + n_samp*3*x):(n_samp*3*x+n_samp), c
```

```r
  # Result for ANOVA for normal residuals and Kruskal-Wallis for non-normal residuals
  n_of_sign_diff = sum(anpv[1,!non_anova_blocks] <0.05) + sum(krpv < 0.05)

  print(paste0("There are significant differences in ", n_of_sign_diff, " pairs of '", varnames[cl-2],
})
```

```
## [1] "There are significant differences in 0 pairs of 'Soil pH at 10cm depth' variable within plots"
## [1] "There are significant differences in 0 pairs of 'Soil temperature, °C' variable within plots"
## [1] "There are significant differences in 0 pairs of 'Active layer thickness, cm' variable within plo
```

As for the "Air temperature, °C" variable, there are equal values of observations within each block, except for the 8th, where temperature of P22 plot is 0.1°C different from two other plots (P23, P24):

```r
ins[ins$block == "BLK08",c(1,2,3)]
```

```
##     block plot air_temp
## 127 BLK08  P22      5.4
## 128 BLK08  P22      5.4
## 129 BLK08  P22      5.4
## 130 BLK08  P22      5.4
## 131 BLK08  P22      5.4
## 133 BLK08  P23      5.5
## 134 BLK08  P23      5.5
## 135 BLK08  P23      5.5
## 136 BLK08  P23      5.5
## 137 BLK08  P23      5.5
## 139 BLK08  P24      5.5
## 140 BLK08  P24      5.5
## 141 BLK08  P24      5.5
## 142 BLK08  P24      5.5
## 143 BLK08  P24      5.5
```

This difference is, of corse, statistically significant, but this slight divation of air temperature is not likely to influence the experiment.

In order to test differences between means of variables in ridge and lakebed locations we perform Mann-Whitney-Wilcoxon test. The diffenences are significant for variables:

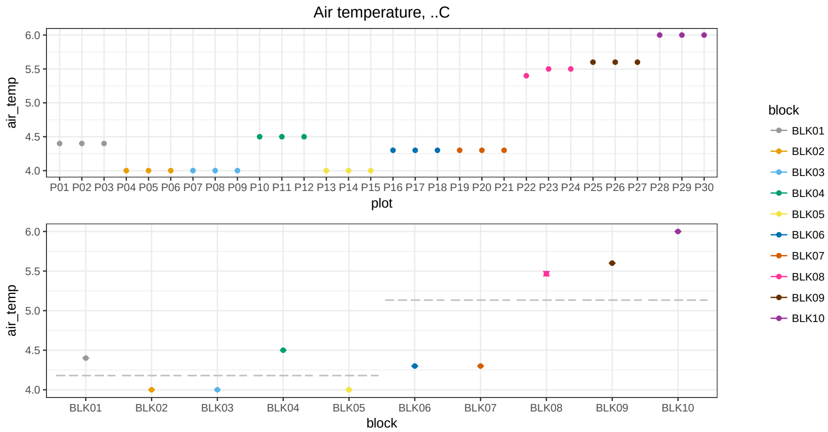
```r
ridge_indexes = ins$block %in% levels(ins$block)[1:5]
mwpv = sapply(c(3:6), function(cl) wilcox.test(ins[ridge_indexes, cl], ins[!ridge_indexes,cl],  exact =
varnames[which(mwpv < 0.05)]
```

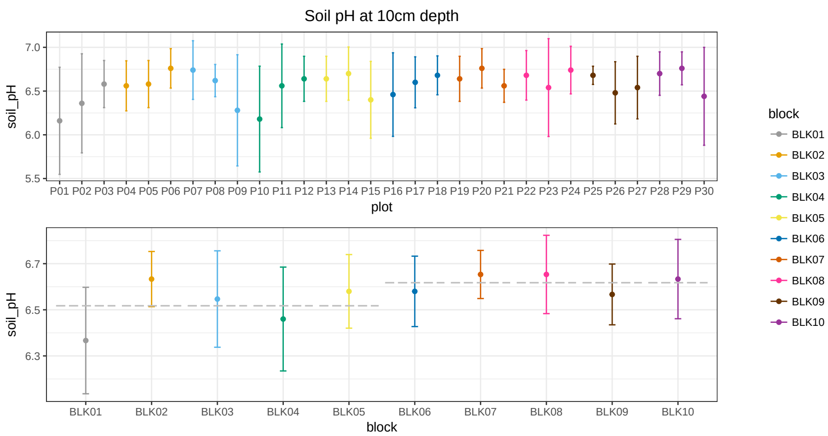
```
## [1] "Air temperature, °C"        "Soil temperature, °C"
## [3] "Active layer thickness, cm"
```

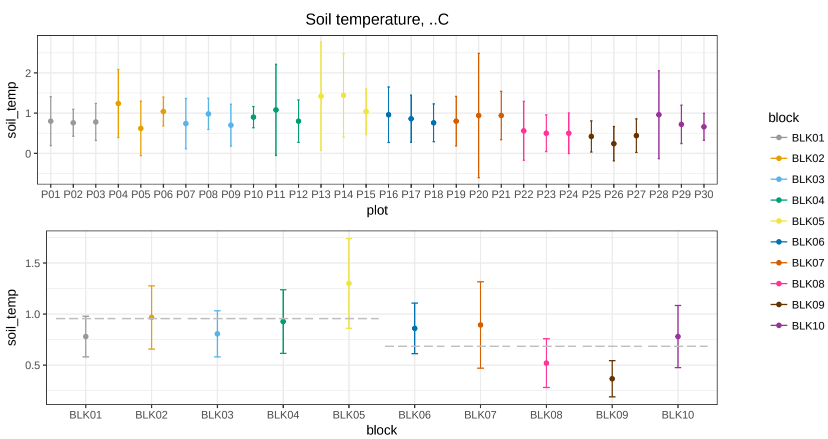
Graphical representation of differences within plots and between plots

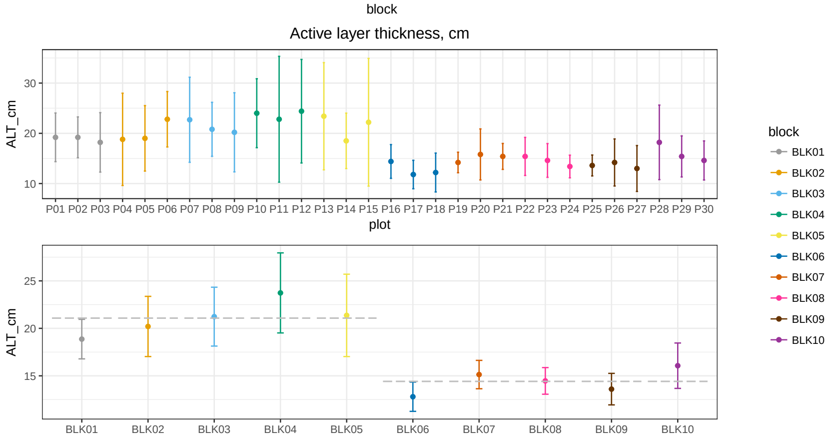
On the figure each dot represents the mean of the variable within plot (top) or block (bottom), vertical lines represent confidance intervals. The top plot for each variable shows variation within each block, while the bottom plot shows variation between blicks. The dashed lines display the means of variables

```r
plot_list <- lapply(c(1:4), function(n) plot_diff(ins, ynames[n], varnames[n]))
cowplot::plot_grid(plotlist=plot_list, ncol=1)
```


Air temperature, ..C
air_temp
plot
block
BLK01
BLK02
BLK03
BLK04
BLK05
BLK06
BLK07
BLK08
BLK09
BLK10


Soil pH at 10cm depth
soil_pH
plot
block
BLK01
BLK02
BLK03
BLK04
BLK05
BLK06
BLK07
BLK08
BLK09
BLK10


Soil temperature, ..C
soil_temp
plot
block
BLK01
BLK02
BLK03
BLK04
BLK05
BLK06
BLK07
BLK08
BLK09
BLK10


Active layer thickness, cm
ALT_cm
plot
block
BLK01
BLK02
BLK03
BLK04
BLK05
BLK06
BLK07
BLK08
BLK09
BLK10

Conclusion

There are no significant differences for soil variables within plots. However, the air temperature has 0.1 difference in one of the plot of 8th block. We also observed significant differences between blocks in ridge and lakebed locations for the "Air temperature, °C", "Soil temperature, °C" and "Active layer thickness, cm" variables.