Appendix. C: Summary of R code used in statistical analyses and visualisation. R version 4.5.1.

Mako Shibata

Preamble

Packages

```
library(tidyr) #formatting
library(dplyr) #manipulation
library(ggplot2) #visualisation
library(readr) #manipulation
library(ggpubr) #visualisation
library(knitr) #visualisation
library(stats) #power analysis
library(pander) #visualisation
library(broom) #visualisation
```

Load Data

```r
# set working directory
setwd("/Users/Owner/Library/CloudStorage/OneDrive-UniversityofEdinburgh/#00_EM/Project_Soil Microclimate/EM_SoilMicroclimate")

# load raw data
soil <- read.csv("RAW_ALL_soil_microclimate.csv")

# combine tree species entry into one column
# remove unnecessary columns
soil <- soil %>%
  mutate(Tree.species = if_else(
    Tree.species == "other",
    Other..specify....Tree.species,
    Tree.species)) %>%
  select(-Other..specify....Tree.species,
         -Mixed.Other..specify....Ground.Cover.Types,
         -Ground.Cover.Types)

# create a long format data, gathered by Distance (0.5 and 7.0 m)
soil0.5 <- soil %>%
  select(Date = Date...Time,
         TreeID = Tree.Plot.ID..,
         Species = Tree.species,
         DBH_cm = DBH..cm.,
         Cardinal = Cardinal.Direction,
         Depth_cm = Soil.organic.layer.depth..cm.,
         Moisture_vv = Soil.moisture....V.V.,
         Distance = Distance.from.tree.trunk..m.,
         Notes = Notes...Comments,
         x = x,
         y = y
         )

soil7.0 <- soil %>%
  select(Date = Date...Time,
         TreeID = Tree.Plot.ID..,
         Species = Tree.species,
         DBH_cm = DBH..cm.,
         Cardinal = Cardinal.Direction,
         Depth_cm = Soil.organic.layer.depth..cm..1,
         Moisture_vv = Soil.moisture....V.V..1,
         Distance = Distance.from.tree.trunk..m..1,
         Notes = Notes...Comments,
         x = x,
         y = y
  )

soil_all <- bind_rows(soil0.5, soil7.0) %>%
  arrange(TreeID, Distance) %>%
  mutate(running_order = row_number()) %>%
  relocate(running_order, .before = TreeID) %>%
  select(-Notes)
```

```
# set distance as a factor
soil_all$Distance <- as.factor(soil_all$Distance)
# omit N/A rows
soil_all <- na.omit(soil_all)
```

Clean data

The following procedure corrects the species mis-entry (English Oak to Silver Birch), adds a column of Genus and management types for group-specific analyses, and creates a new data frame with no Scots Pine data. Scots Pine is removed due to its sample size being 1.

```
# replace oak with birch, set genus name in a new column.
soil_all <- soil_all %>%
  mutate(Species = recode(Species,
                          "English Oak (Quercus robur)" = "Silver Birch (Betula pendu
la)")) %>%
  mutate(Genus = case_when(
    Species == "Alder" ~ "Alnus",
    Species == "Silver Birch (Betula pendula)" ~ "Betula",
    Species == "Downy Birch (Betula pubescens)" ~ "Betula",
    Species == "Scots pine (Pinus sylvestris)" ~ "Pinus",
    Species == "Rowan (Sorbus aucuparia)" ~ "Sorbus"
  ))  %>%
  relocate(Genus, .before = Species)

# create a management type column using TreeID that falls within a specific managemen
t regime
soil_all<- soil_all %>%
  mutate(soil_all, Management = case_when(
    TreeID %in% c(1, 2, 3, 4, 5, 6, 7, 21, 22, 23, 24, 25) ~ "No mound",
    TRUE ~ "mound")) %>%
  relocate(Management, .after = Species)

# save cleaned data including pine
write.csv(soil_all, "soil_all.csv")

# remove pine
soil_all_nonpine <- soil_all %>%
  filter(Genus != "Pinus")

# save cleaned data excluding pine
write.csv(soil_all_nonpine, "soil_all_nonpine.csv")
```

DBH (cm) and Age

The following codes calculated mean, standard deviation, max, and minimum of DBH (cm) for total species and all genus groups. DBH is converted to age by applying a growth factor of 2.5 cm/year (Mitchell, 1974).

```r
# read soil_all_pine as soil
soil <- soil_all_nonpine
# get average DBH (cm) +- std for the following groups: total, alder, birch, sorbus
DBH_mean <- c(mean(soil$DBH_cm),
              mean(soil$DBH_cm[soil$Genus == "Alnus"]),
              mean(soil$DBH_cm[soil$Genus == "Betula"]),
              mean(soil$DBH_cm[soil$Genus == "Sorbus"])))

DBH_sd <- c(sd(soil$DBH_cm), sd(soil$DBH_cm[soil$Genus == "Alnus"]),
            sd(soil$DBH_cm[soil$Genus == "Betula"]),
            sd(soil$DBH_cm[soil$Genus == "Sorbus"])))

DBH_max <- c(max(soil$DBH_cm),
             max(soil$DBH_cm[soil$Genus == "Alnus"]),
             max(soil$DBH_cm[soil$Genus == "Betula"]),
             max(soil$DBH_cm[soil$Genus == "Sorbus"])))

DBH_min <- c(min(soil$DBH_cm),
             min(soil$DBH_cm[soil$Genus == "Alnus"]),
             min(soil$DBH_cm[soil$Genus == "Betula"]),
             min(soil$DBH_cm[soil$Genus == "Sorbus"])))

Genus <- c("Total", "Alnus", "Betula", "Sorbus")

# create a data frame
DBH_Age <- data.frame(Genus, DBH_mean, DBH_sd, DBH_max, DBH_min)
names(DBH_Age) <- c("Genus", "meanDBH", "stdDBH", "maxDBH", "minDBH")

# converting this to girth
DBH_Age <- DBH_Age %>%
  mutate(girth_cm = meanDBH*3.14,
         sd_girth_cm = stdDBH*3.14,
         girth_max = DBH_max*3.14,
         girth_min = DBH_min*3.14)

# applying the growth rate = 2.5 cm to estimate age
DBH_Age<- DBH_Age %>%
  mutate(mean_age = girth_cm/2.5,
         age_sd = sd_girth_cm/2.5,
         maxage = girth_max/2.5,
         minage = girth_min/2.5)

write.csv(DBH_Age, "DBH_Age.csv")

# outputs
DBH_Age_outputs <- DBH_Age %>%
  select(Genus, mean_age, age_sd, maxage, minage)
kable(DBH_Age_outputs, digits = 1)
```

Genus	mean_age	age_sd	maxage	minage
Total	10.3	6.8	32.7	4.0
Alnus	10.7	10.9	28.4	4.0
Betula	10.5	6.1	32.7	4.3

Genus	mean_age	age_sd	maxage	minage
Sorbus	8.6	3.6	12.8	4.9

Power Analyses

Code

The following addresses the question; what is the sample size required to run one-way ANOVA?

```
# One-Way ANOVA setting
# alpha = 0.05
# power = 0.80
# group = 2

power.anova.test(groups = 2, between.var = 1, within.var = 3, power = 0.8)
```

```
##
##       Balanced one-way analysis of variance power calculation
##
##          groups = 2
##               n = 24.54495
##      between.var = 1
##       within.var = 3
##        sig.level = 0.05
##            power = 0.8
##
## NOTE: n is number in each group
```

```
## n = 25 per group for between.var = 1, within.var = 3
```

Power Analyses Results, Statistical Test Selection

The power analysis above indicated that the total sample size (n = 24, excluding Scots pine) was sufficient to detect large effect sizes (f = 0.5) in one-way ANOVA for the full data set (all trees).

Power analyses for two-way ANOVA were done using G*Power assuming an effect size of 0.5 and power of 0.8. This resulted in an even sample size of 10 per category as a requirement to run two-way ANOVA. Using two-way ANOVA in this case for interaction effects of management and genus can lead to misinterpretation. To avoid this, we opted to perform one-way ANOVA for each factor separately.

Visualisation: All tree & Genus specific

```r
# mutate "total" category in Genus column
datavis <- soil_all_nonpine %>%
  mutate(Genus = "Total")
datavis <- bind_rows(
  soil_all_nonpine, datavis
)

datavis$Genus <- factor(datavis$Genus, levels = c("Total", "Alnus", "Betula", "Sorbus"))

p1 <- ggbarplot(
  datavis, x = "Distance", y = "Depth_cm",
  add = c("mean_se", "jitter"),
  add.params = list(shape = "Genus"),
  fill= "Genus", palette = c(
    "Total" = "#8B4513",
    "Alnus" = "#E0EEEE",
    "Betula" = "#C1CDCD",
    "Sorbus" = "#838B8B"),
  position = position_dodge(0.8)
) +
  labs(x = "Distance from a tree trunk (m)", y = "O Horizon Depth (cm)")

p2 <- ggbarplot(
  datavis, x = "Distance", y = "Moisture_vv",
  add = c("mean_se", "jitter"),
  add.params = list(shape = "Genus"),
  fill= "Genus", palette = c(
    "Total" = "#6495ED",
    "Alnus" = "#E0EEEE",
    "Betula" = "#C1CDCD",
    "Sorbus" = "#838B8B"),
  position = position_dodge(0.8)
) +
  labs(x = "Distance from a tree trunk (m)", y = "Soil moisture (%, v/v)")

ggarrange(p1, p2,
          nrow = 1, ncol = 2,
          labels = c("a", "b"))
```
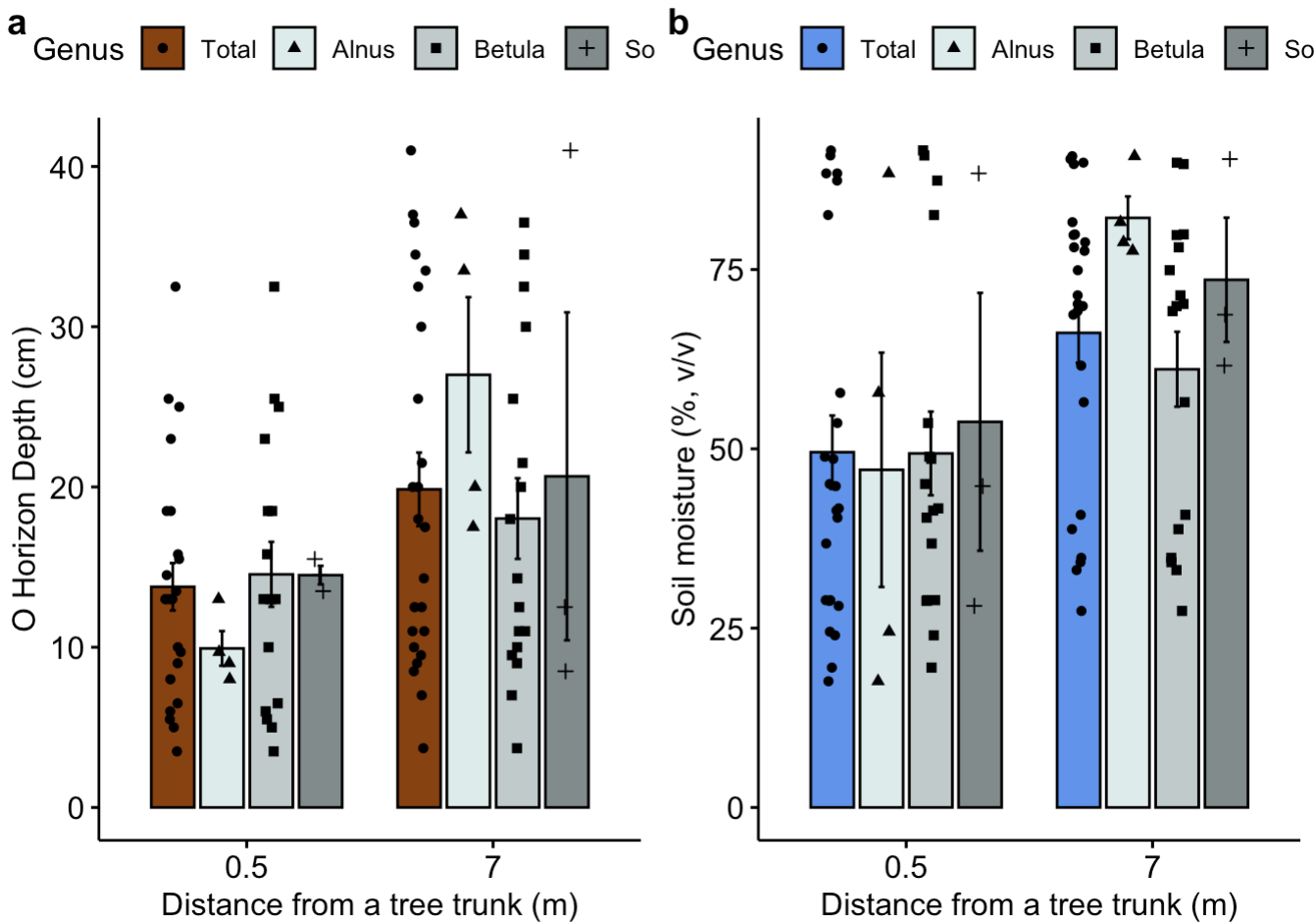
```
dev.off()
```

```
## null device
##           1
```

Linear model: ANOVA for All tree & Genus specific

```
# All trees
## ANOVA Depth vs. Distance----
depth_lm = lm(Depth_cm ~ Distance, data = soil_all_nonpine)
depth_anov = aov(depth_lm, data = soil_all_nonpine)

table_obj <- tidy(depth_anov)
pander(depth_anov, digits = 3)
```
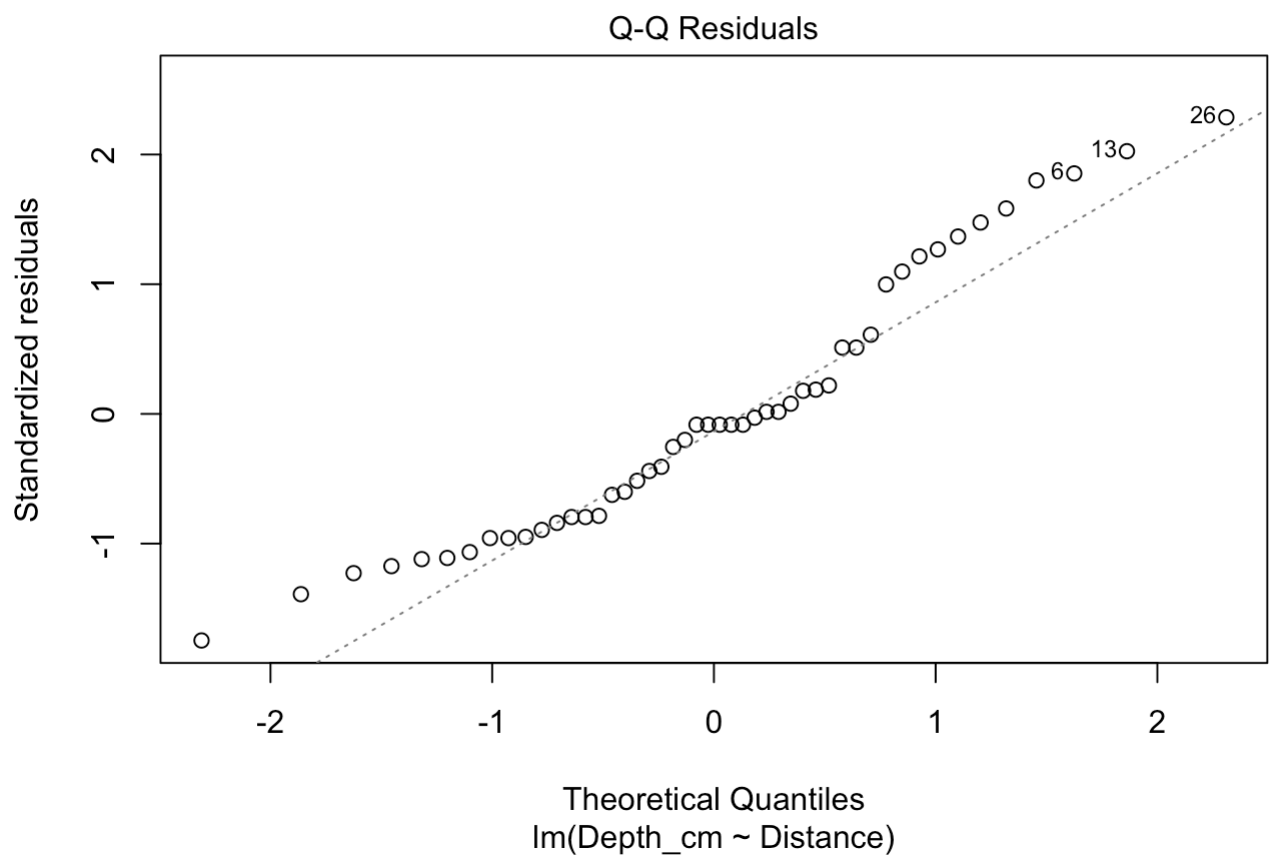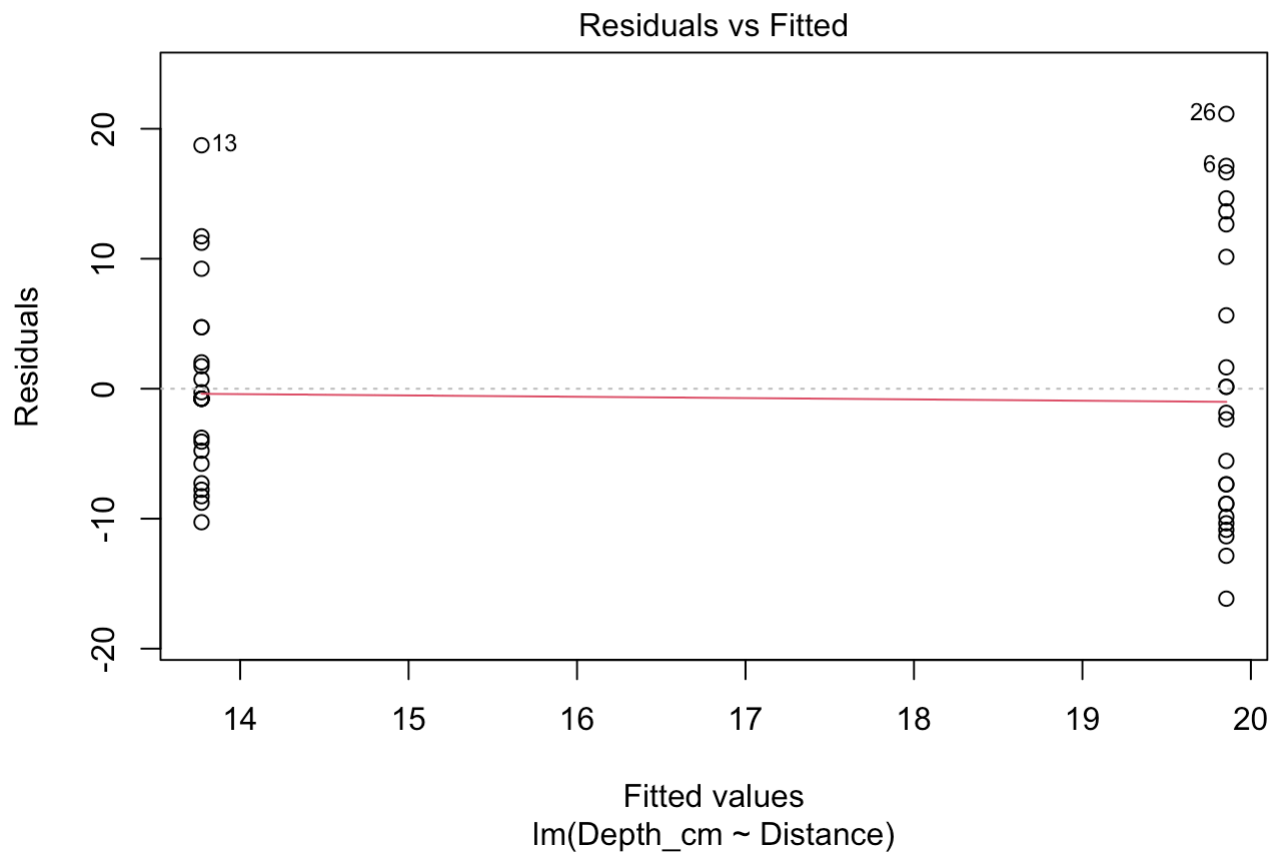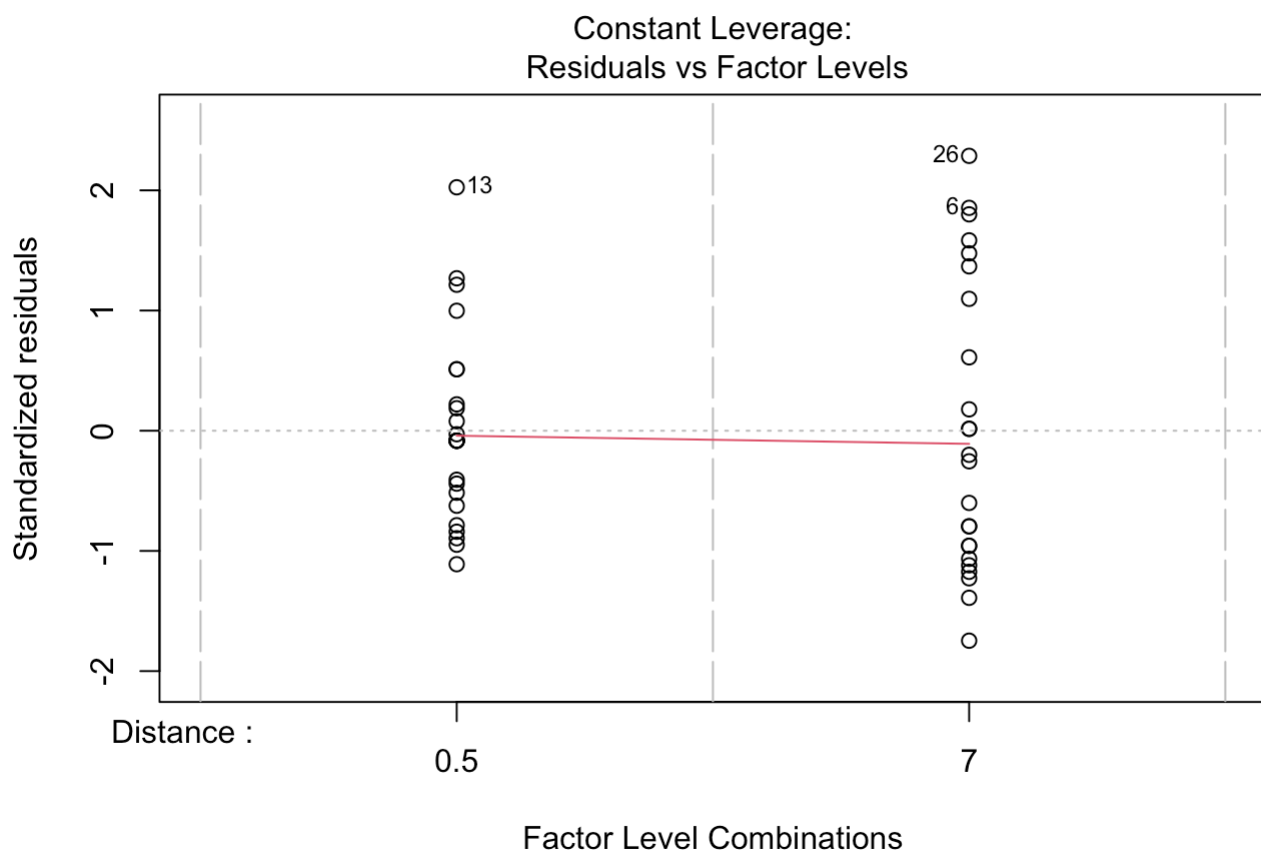
Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	444	444	4.98	0.0306
Residuals	46	4104	89.2	NA	NA

```
## calculating mean
tapply(soil_all_nonpine$Depth_cm, soil_all_nonpine$Distance, mean, na.rm=TRUE)
```

```
##       0.5        7
## 13.77083 19.85417
```

```
## Anscombe's Quartet
plot(depth_lm)
```

Residuals vs Fitted

lm(Depth_cm ~ Distance)

Q-Q Residuals

lm(Depth_cm ~ Distance)

```
# Residuals vs. fitted: linearity met.
# Q-Q plot generally follows the same line, residuals normally distributed.
# Residuals variances are equal.
# No significant outliers outside the Cook's distance.

## ANOVA Moisture vs. Distance----
moisture_lm <- lm(Moisture_vv ~ Distance, data = soil_all_nonpine)
moisture_anov <- aov(moisture_lm)

table_obj <- tidy(moisture_anov)
pander(moisture_anov, digits = 3)
```
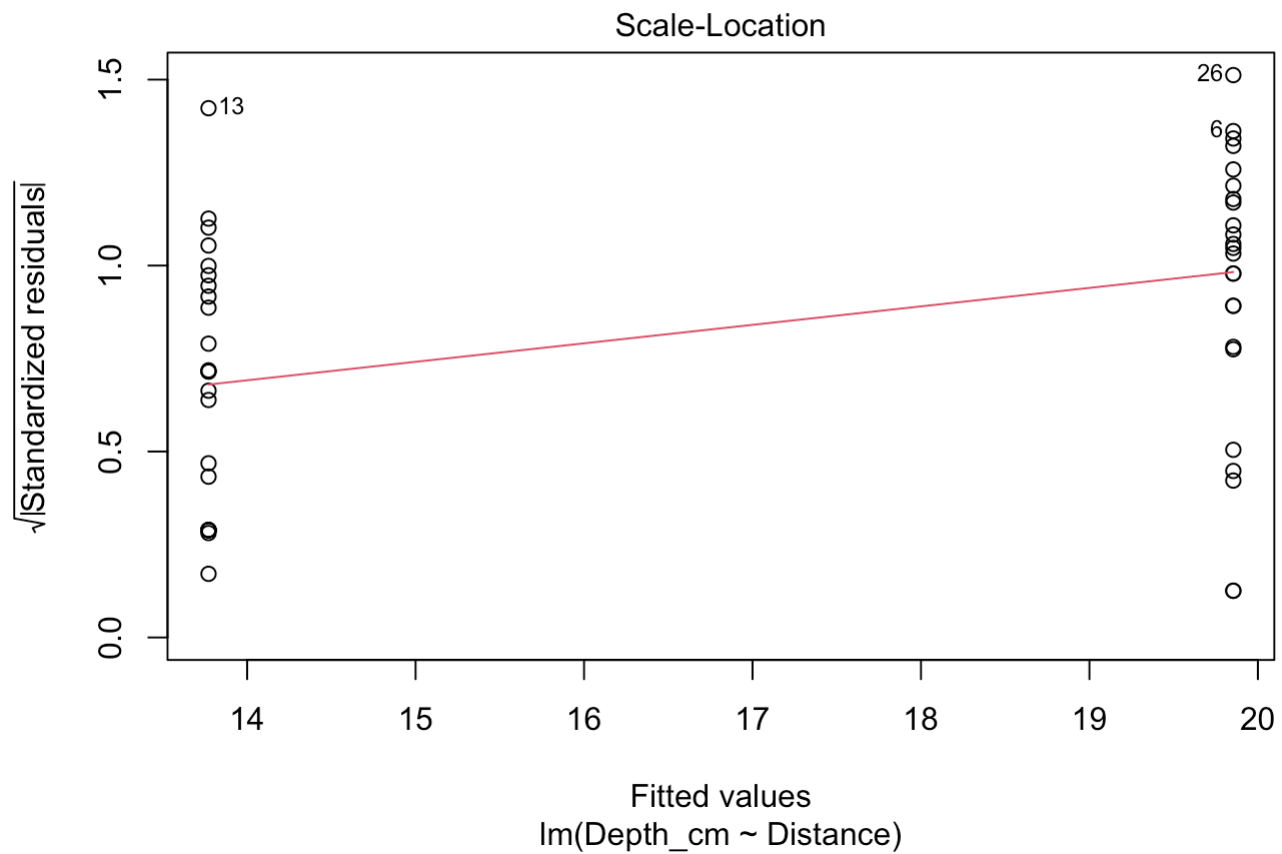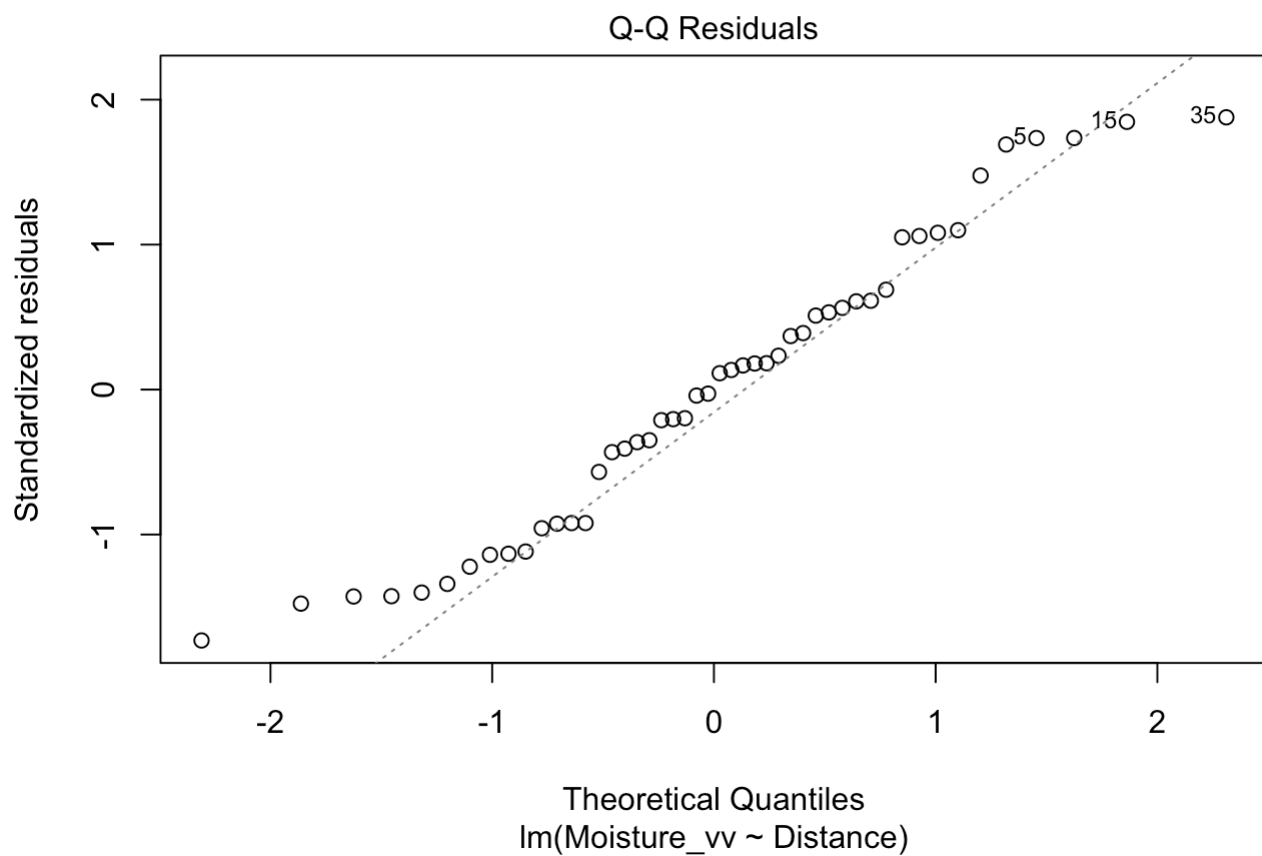
Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	3323	3323	6.35	0.0153
Residuals	46	24086	524	NA	NA

```
## Anscombe's Quartet
plot(moisture_lm)
```

Residuals vs Fitted

lm(Moisture_vv ~ Distance)

Q-Q Residuals

lm(Moisture_vv ~ Distance)

Scale-Location

lm(Moisture_vv ~ Distance)

Constant Leverage:
Residuals vs Factor Levels

Factor Level Combinations
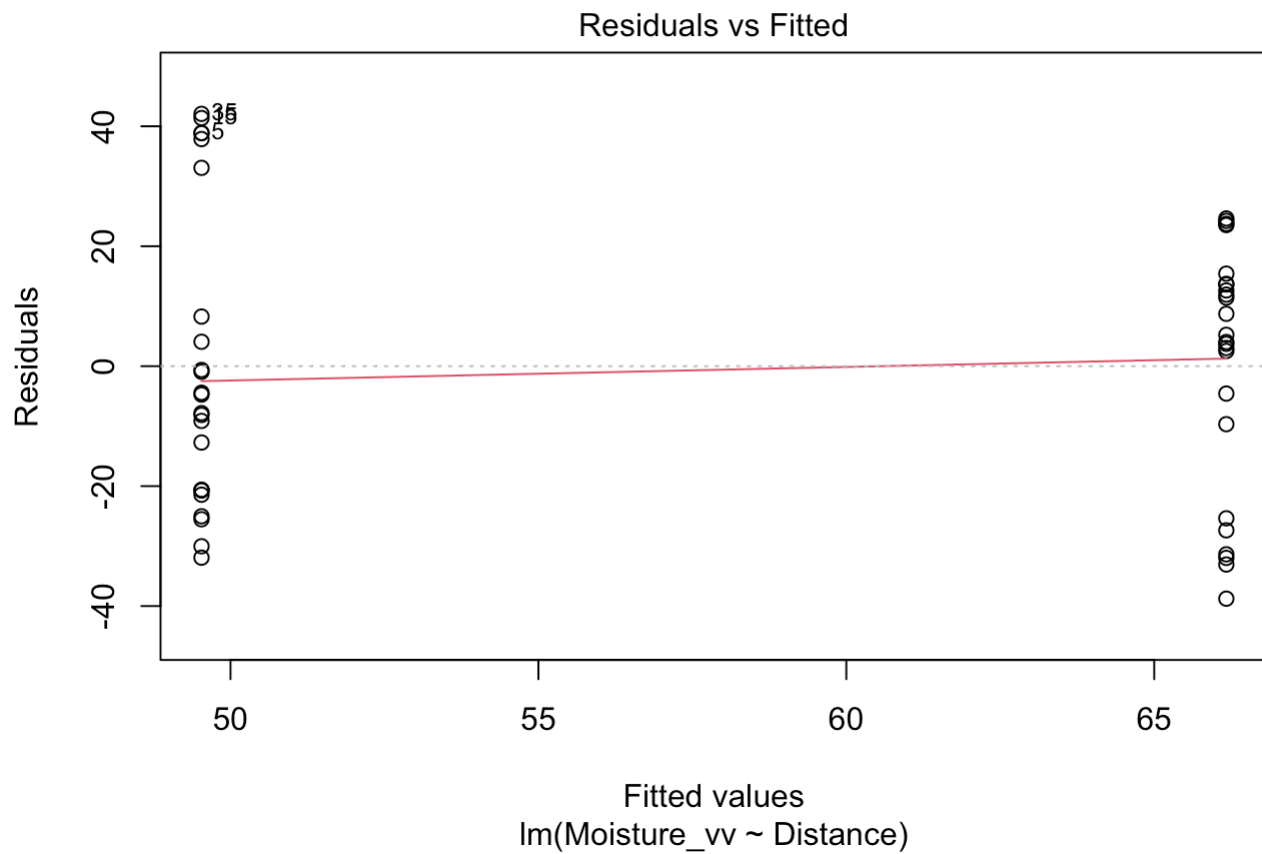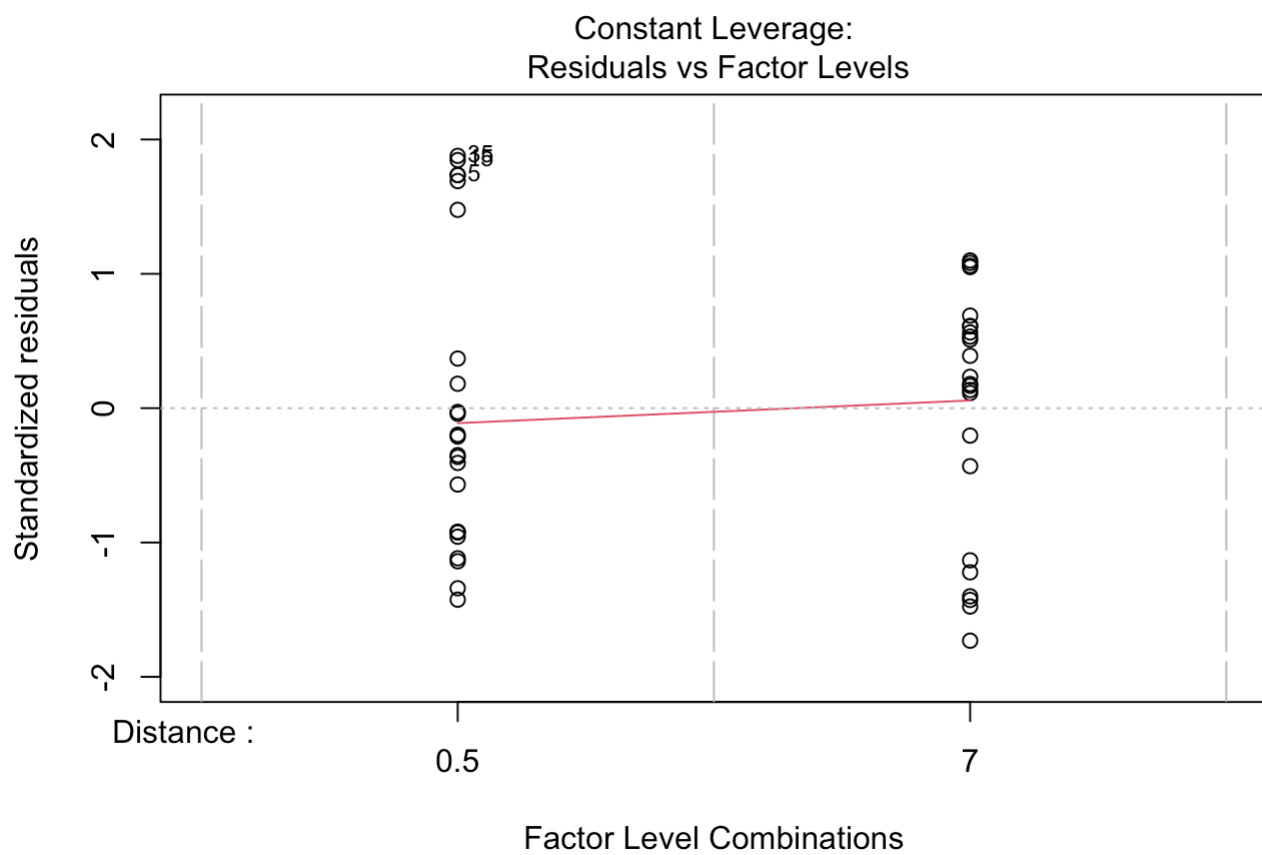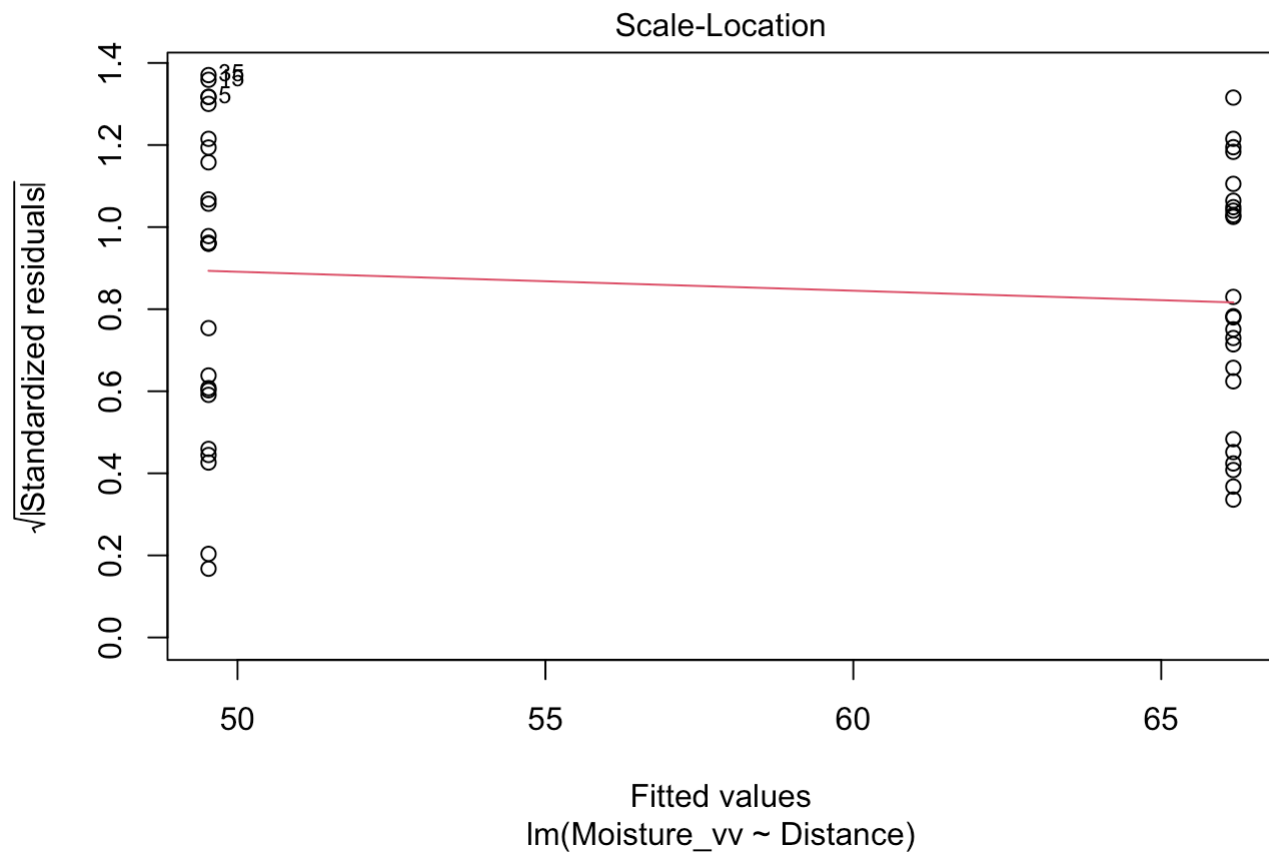
```r
# Linearity met.
# Q-Q plot generally follows the same line.
# Residual variances are generally equal.
# No significant outliers outside the Cook's distance.

# Species Specific
## ANOVA Depth vs. Distance----
## Subset data for One-way ANOVA
Alnus <- subset(soil_all_nonpine,Genus == "Alnus")
Betula <- subset(soil_all_nonpine, Genus == "Betula")
Sorbus <- subset(soil_all_nonpine, Genus == "Sorbus")

## ANOVA
alnus_mod <- aov(Depth_cm ~ Distance, data = Alnus)
table_obj <- tidy(alnus_mod)
pander(alnus_mod, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	583	583	11.8	0.0138
Residuals	6	296	49.3	NA	NA

```r
# p = 0.01, df = 6. Significant difference.
betula_mod <- aov(Depth_cm ~ Distance, data = Betula)
table_obj <- tidy(betula_mod)
pander(betula_mod, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	103	103	1.16	0.29
Residuals	32	2845	88.9	NA	NA

```r
# p = 0.29, df = 32, not significant.
sorbus_mod <- aov(Depth_cm ~ Distance, data = Sorbus)
table_obj <- tidy(sorbus_mod)
pander(sorbus_mod, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	57	57	0.362	0.58
Residuals	4	630	158	NA	NA

```
# p = 0.58, df = 4, not significant.

## ANOVA Moisture vs. Distance----

## ANOVA
alnus_mod <- aov(Moisture_vv ~ Distance, data = Alnus)
table_obj <- tidy(alnus_mod)
pander(alnus_mod, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	2468	2468	4.48	0.0788
Residuals	6	3308	551	NA	NA

```
# p = 0.08, df = 6. not significant.
betula_mod <- aov(Moisture_vv ~ Distance, data = Betula)
table_obj <- tidy(betula_mod)
pander(betula_mod, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	1171	1171	2.25	0.144
Residuals	32	16681	521	NA	NA

```
# p = 0.14, df = 32, not significant.
sorbus_mod <- aov(Moisture_vv ~ Distance, data = Sorbus)
table_obj <- tidy(sorbus_mod)
pander(sorbus_mod, digits = 3)
```

Analysis of Variance Model

|	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	588	588	0.985	0.377
Residuals	4	2389	597	NA	NA

```
# p = 0.38, df = 4, not significant.
```

Visualisation: Management specific

```
## Distance vs O Horizon depth (cm)
p3 <- ggbarplot(
  soil_all_nonpine, x = "Distance", y = "Depth_cm",
  add = c("mean_se", "jitter"),
  add.params = list(shape = "Management"),
  fill= "Management", palette = c(
    "No mound" = "#D9C2C1",
    "mound" = "gray"),
  position = position_dodge(0.8)
) +
  labs(x = "Distance from a tree trunk (m)", y = "O Horizon Depth (cm)")

## Distance vs Soil moisture (cm)
p4 <- ggbarplot(
  soil_all_nonpine, x = "Distance", y = "Moisture_vv",
  add = c("mean_se", "jitter"),
  add.params = list(shape = "Management"),
  fill= "Management", palette = c(
    "No mound" = "#A7B9E8",
    "mound" = "gray"),
  position = position_dodge(0.8)
) +
  labs(x = "Distance from a tree trunk (m)", y = "Soil moisture (%, v/v)")

ggarrange(
  p3, p4,
  labels = c("a", "b"),
  ncol = 2, nrow = 1,
  common.legend = FALSE
)
```
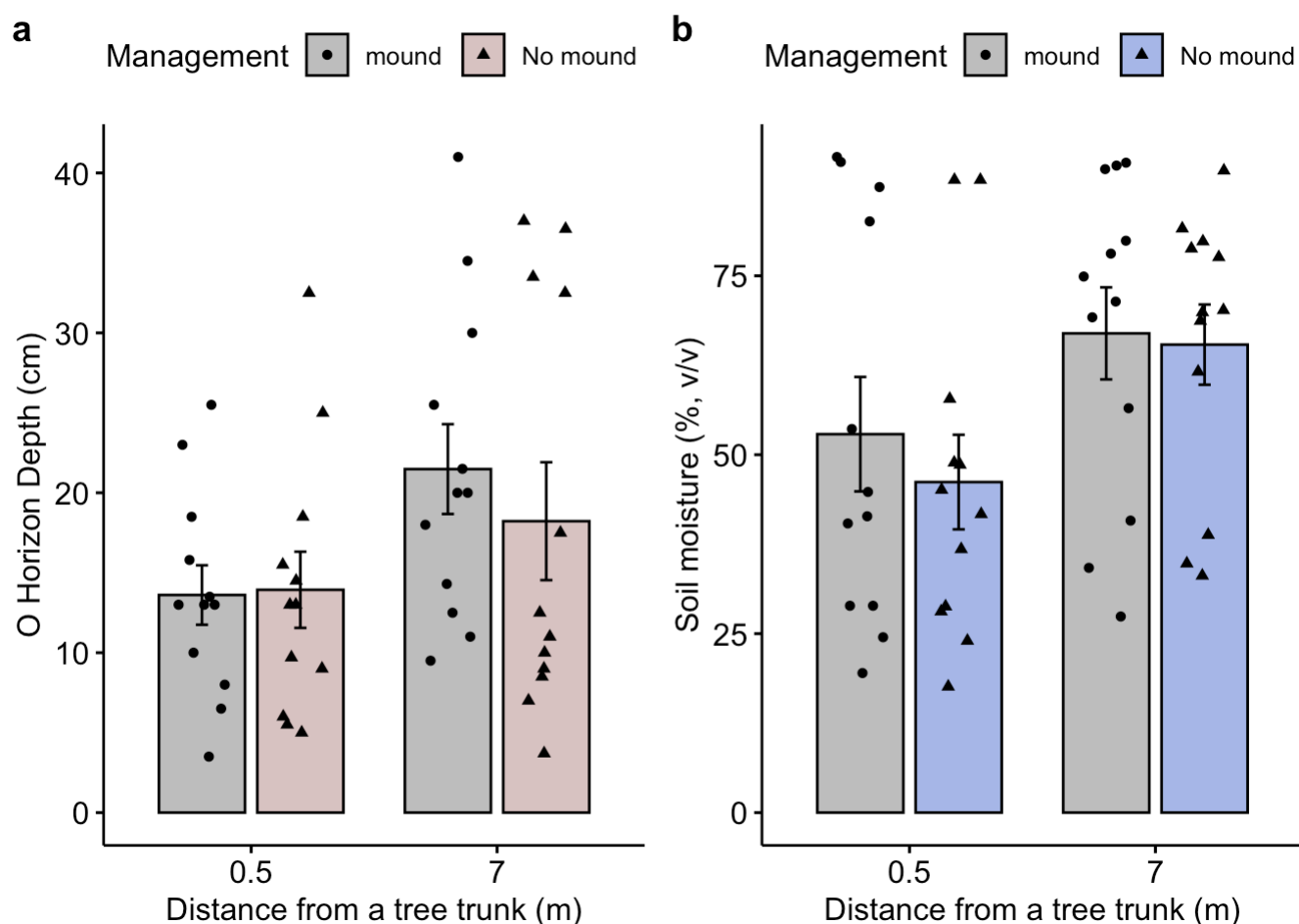
Linear model: ANOVA for Management specific

```
# Subset data
no_mound <- soil_all_nonpine %>%
  subset(Management == "No mound")
mound <- soil_all_nonpine %>%
  subset(Management == "mound")

# ANOVA
## Depth vs. Distance ----
### Mound
mound_aov <- aov(Depth_cm ~ Distance, data = mound)
table_obj <- tidy(mound_aov)
pander(mound_aov, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	372	372	5.46	0.0289
Residuals	22	1498	68.1	NA	NA

```
### No mound
no_mound_aov <- aov(Depth_cm ~ Distance, data = no_mound)
table_obj <- tidy(no_mound_aov)
pander(no_mound_aov, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	111	111	0.956	0.339
Residuals	22	2542	116	NA	NA

```
# p = 0.05 for mound, p = 0.34 for no-mound.
# df = 24, and df = 24, respectively.

## Moisture vs. Distance ----
### Mound
mound_aov <- aov(Moisture_vv ~ Distance, data = mound)
table_obj <- tidy(mound_aov)
pander(mound_aov, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	1190	1190	1.88	0.184
Residuals	22	13898	632	NA	NA

```
### No mound
no_mound_aov <- aov(Moisture_vv ~ Distance, data = no_mound)
table_obj <- tidy(no_mound_aov)
pander(no_mound_aov, digits = 3)
```

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distance	1	2212	2212	4.91	0.0373
Residuals	22	9905	450	NA	NA

```
# p = 0.22 for mound, p = 0.04 for no-mound.
# df = 24 and df = 24 respectively.
```

Means ± SEs for ALL Groups

```r
# total trees
total_summary <- soil_all_nonpine %>%
  group_by(Distance) %>%
   summarise(
    Mean_depth = mean(Depth_cm, na.rm = TRUE),
    SE_depth  = sd(Depth_cm, na.rm = TRUE) / sqrt(n()),
    Mean_moist = mean(Moisture_vv, na.rm = TRUE),
    SE_moist  = sd(Moisture_vv, na.rm = TRUE) / sqrt(n())
  )
# grouped by genus
genus_summary <- soil_all_nonpine %>%
  group_by(Genus, Distance) %>%
  summarise(
    Mean_depth = mean(Depth_cm, na.rm = TRUE),
    SE_depth  = sd(Depth_cm, na.rm = TRUE) / sqrt(n()),
    Mean_moist = mean(Moisture_vv, na.rm = TRUE),
    SE_moist  = sd(Moisture_vv, na.rm = TRUE) / sqrt(n())
  )

# grouped by management
management_summary <- soil_all_nonpine %>%
  group_by(Management, Distance) %>%
  summarise(
    Mean_depth = mean(Depth_cm, na.rm = TRUE),
    SE_depth  = sd(Depth_cm, na.rm = TRUE) / sqrt(n()),
    Mean_moist = mean(Moisture_vv, na.rm = TRUE),
    SE_moist  = sd(Moisture_vv, na.rm = TRUE) / sqrt(n())

  )

summary_table <- bind_rows(
  total_summary %>%
    mutate(Group = "Overall", Category = "All"),
  genus_summary %>%
    mutate(Group = "Genus") %>%
    rename(Category = Genus),
  management_summary %>%
    mutate(Group = "Management") %>%
    rename(Category = Management)
) %>%
  select(Group, Category, Distance, Mean_depth, SE_depth, Mean_moist, SE_moist, Dista
nce)

kable(summary_table, digits = 2, caption = "Summary of O Horizon Depth (cm) and Soil
Moisture (%) (Mean ± SE)")
```

Summary of O Horizon Depth (cm) and Soil Moisture (%) (Mean ± SE)

Group	Category	Distance	Mean_depth	SE_depth	Mean_moist	SE_moist
Overall	All	0.5	13.77	1.48	49.53	5.12
Overall	All	7	19.85	2.29	66.17	4.17
Genus	Alnus	0.5	9.93	1.08	47.08	16.33

Group	Category	Distance	Mean_depth	SE_depth	Mean_moist	SE_moist
Genus	Alnus	7	27.00	4.84	82.20	2.99
Genus	Betula	0.5	14.55	2.03	49.36	5.83
Genus	Betula	7	18.03	2.52	61.09	5.23
Genus	Sorbus	0.5	14.50	0.58	53.77	17.98
Genus	Sorbus	7	20.67	10.23	73.57	8.66
Management	No mound	0.5	13.93	2.38	46.18	6.61
Management	No mound	7	18.23	3.69	65.38	5.60
Management	mound	0.5	13.61	1.86	52.88	8.00
Management	mound	7	21.48	2.81	66.96	6.43

References

Mitchell, A.F. (1974) A field guide to the trees of Britain and northern Europe. Boston : Houghton Mifflin.
Available at: http://archive.org/details/fieldguidetotree00mitc (http://archive.org/details/fieldguidetotree00mitc)
(Accessed: 3 October 2025).