Variability in tree height and diameter across select species in Alaska, USA

Vignesh Arunkumar, Aidan Power, Brynn Rotbart, Sophie Valkenberg

2024-12-03

Contents

List of Figures

Setting up our R Workspace

```
# Set your working directory
getwd()
```

```
## [1] "/home/guest/EDEProj_Arunkumar_Power_Rotbart_Valkenberg"
```

```
setwd("/home/guest/EDEProj_Arunkumar_Power_Rotbart_Valkenberg")
```

```
# Load your datasets
Ak_data <-
  read.csv(
    '~/EDEProj_Arunkumar_Power_Rotbart_Valkenberg/Data_Overall/Raw/AK_SITETREE.csv')
```

Rationale and Research Questions

The topic of tree heights was chosen because it is one of the main things that our group has in common; two of us are in the forestry program, while the other two are TFE concentrations. In addition, we all found a similar interest in discovering not only how tree heights vary with tree diameter and species, but also exploring these factors in a region unknown to us.

Alaska is home to a lush and vibrant ecosystem–one that is entirely foreign to North Carolina. Because of the freedom of this project, we wanted to take the opportunity to learn more about an entirely different variation of flora that was unknown to us. Our ignorance, combined with the dataset, including height, DBH, and species codes, motivated us to use this dataset to answer our questions. Additionally, the data comes from the United States Forest Service, which is a group we all have learned much about and wanted to interact with. This leads us to the main question of this research:

Q: How does tree height differ among specific tree species in Alaska, including white spruce, black spruce, lodgepole pine, and mountain hemlock?

H0:The mean tree species height does not vary enough to be significant

H1: The mean tree height differs significantly across species

Following questions to help us further understand the differences in these species would be:

- How does this height vary with DBH (diameter at breast height)?
- How does this height vary with species?
- How does this height vary with plot?

Data Wrangling

```
#Filtering for wanted columns
Ak_data.wrangled <- Ak_data %>%
  select(PLOT, SPCD, DIA, HT)

#Filtering for Species
```

```
Species.wanted <- Ak_data.wrangled %>%
  filter(SPCD %in% c(94, 95, 108, 264))
Species.wanted <- Species.wanted %>%
  mutate(Species = case_when(
    SPCD == 94 ~ "White Spruce",
    SPCD == 95 ~ "Black Spruce",
    SPCD == 108 ~ "Lodgepole Pine",
    SPCD == 264 ~ "Mountain Hemlock"
  ))
Species.wanted$Species <- as.factor(Species.wanted$Species)
Species.wanted$PLOT <- as.factor(Species.wanted$PLOT)

write.csv(Species.wanted,
file =
  '~/EDEProj_Arunkumar_Power_Rotbart_Valkenberg/Data_Overall/Processed/AK_SITETREE_SpeciesWanted.csv',
row.names = FALSE)
```

GGPlot Theme Setup

```
# Set your ggplot theme

our_theme <- theme(
  axis.text = element_text(color = "white"),
  legend.position = "top",
  plot.background = element_rect(fill = "#006400",color = NA),
  panel.background = element_rect(fill = "#b2e6b2", color = NA),
  plot.title = element_text(face = "bold", color="white", hjust = 0.5),
  plot.subtitle = element_text(face = "bold", color="white", hjust = 0.5),
  axis.title.x = element_text(color="white"),
  axis.title.y = element_text(angle=90, color="white")
)
theme_set(our_theme)
```

Dataset Information

Data was collected using both remote sensing and ground sampling. Remote sensing was used to group the trees into classes based on similar strata, and used stratum weight as well as known total area to estimate population totals. Ground sampling was done via plots to cover a one-acre sample area, with those plots either being new or re-measurements of old plots. Among the variables measured were tree diameter, height, and species, all three being very important to our research questions and subsequent analysis.

Once we downloaded the dataset, our process of data wrangling involved differentiating the trees by species. We placed a special emphasis on this due to how the dataset only provides species codes, not the the names of the species themselves. Therefore, we filtered for the four most common species, changing their codes into their species names in a new separate column. We also turned everything in this species column to a factor in order to for it to fit in the analysis.

`Table: Terms and Units for Alaska SITETREE Dataset`

Item Name	Value
INVYR	Year of inventory
SPCD	Species code
DIA	Diameter at breast height (in)
HT	Total height (ft)
AGEDIA	Tree age at diameter (years)
METHOD	Method for determining site index (1:collected this inventory 2:collected last inventory 3:estimated 4:height-intercept method this inventory)
SITREE_FVS	Site index of tree (height that tree is expected to attain at reference age)
SIBASE_FVS	Site index base age (Set in years to the closest rotation/culmination year of mean annual increment)

Exploratory Analysis

Below, we've included the key exploratory analysis steps that allowed us to determine which steps should follow in order to answer our research question.

Linear Regressions

Question 1: How does height vary with DBH?

Correlation between Height and Diameter for all Species

```
##
## Call:
## lm(formula = HT ~ DIA, data = Ak_data.wrangled)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -111.375   -8.166   -1.053    7.420   68.400
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.07053    0.21767   50.86   <2e-16 ***
## DIA          3.77365    0.01703  221.54   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.62 on 18598 degrees of freedom
##   (89 observations deleted due to missingness)
## Multiple R-squared:  0.7252, Adjusted R-squared:  0.7252
## F-statistic: 4.908e+04 on 1 and 18598 DF,  p-value: < 2.2e-16
```

This first linear regression that is run shows the correlation between height and diameter at breast height for all species and trees sampled within the data set. Looking at the results it is clear there is a direct positive correlation between height and diameter. As one increases the other does as well.

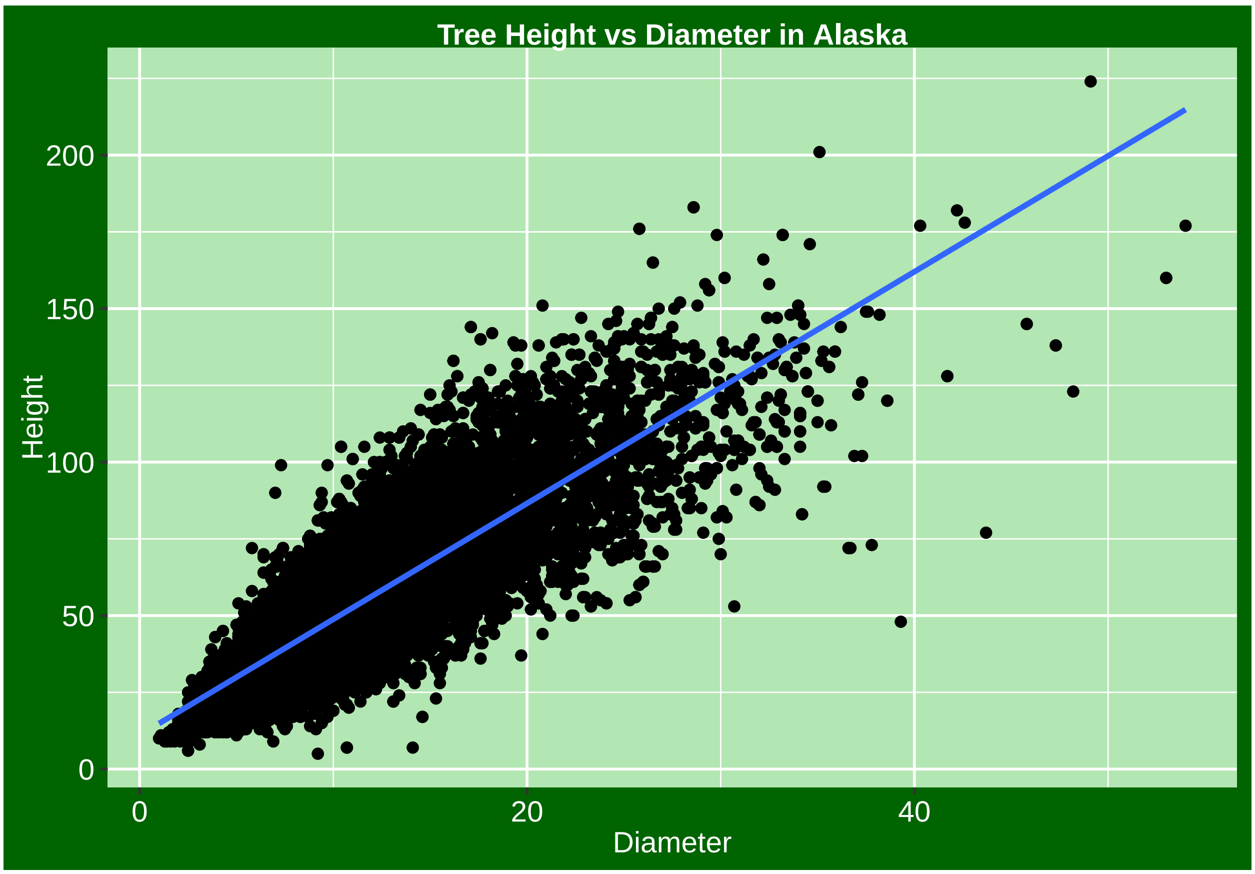

Figure 1: This figure displays the correlation between height and diameter for all species

Question 2: How does height vary with species?

Correlation between Height and Diameter taking Species of Interest into Account

```
##
## Call:
## lm(formula = HT ~ DIA + Species, data = Species.wanted)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.844  -5.147  -0.709   5.008  43.510
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             11.22077    0.23518   47.71   <2e-16 ***
## DIA                      3.65812    0.02952  123.92   <2e-16 ***
## SpeciesLodgepole Pine   -9.17609    0.37384  -24.55   <2e-16 ***
## SpeciesMountain Hemlock -9.14099    0.31694  -28.84   <2e-16 ***
## SpeciesWhite Spruce      2.96974    0.27624   10.75   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.691 on 8534 degrees of freedom
##   (82 observations deleted due to missingness)
## Multiple R-squared:  0.7087, Adjusted R-squared:  0.7086
## F-statistic:  5192 on 4 and 8534 DF,  p-value: < 2.2e-16
```

Now going specifically into chosen species. We wanted to see the correlation between height and dbh for four specific species, Black Spruce, Lodgepole Pine, Mountain Hemlock, White Spruce. From our graph we are able to see the different correlations between each species. With Black spruce's height being less effected by Diameter than the other species.

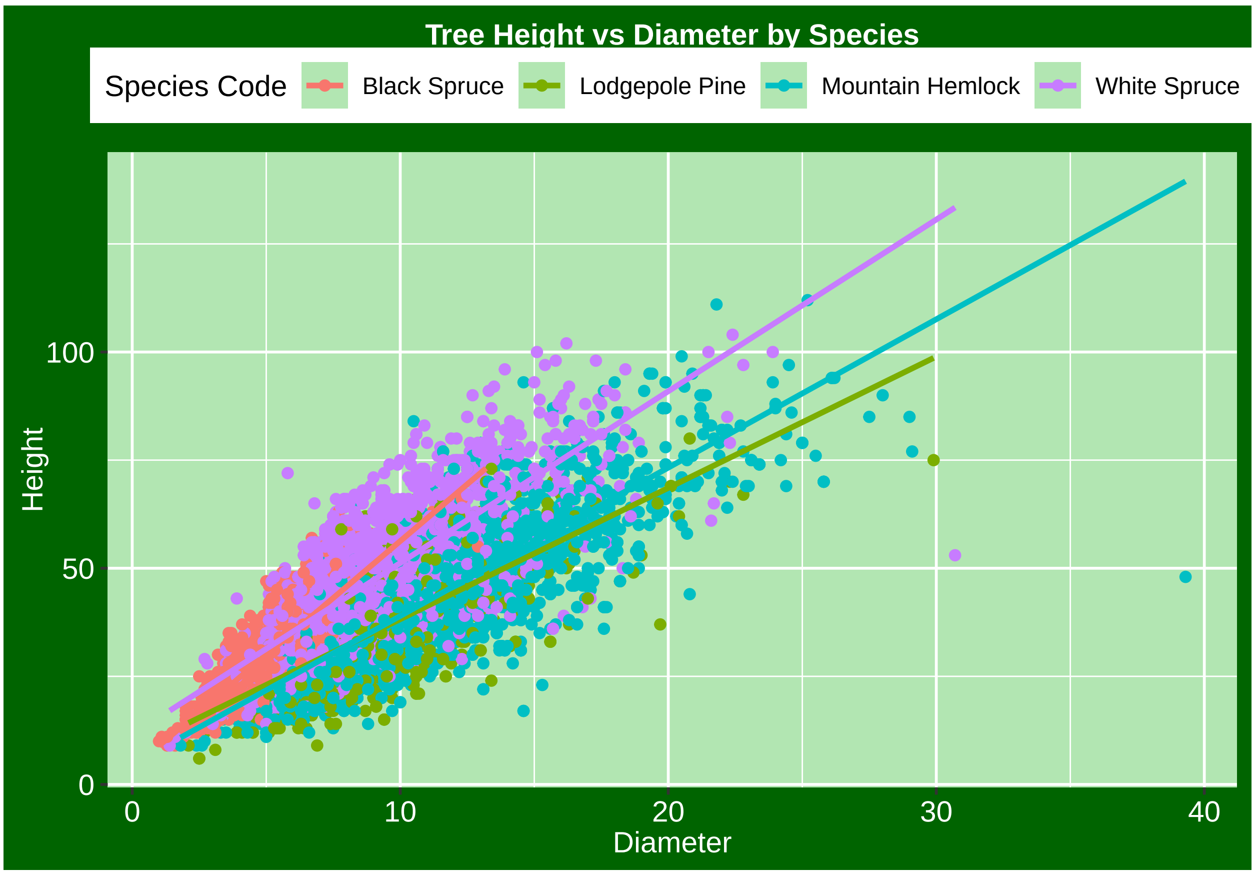

Figure 2: This figure displays the correlation between height and diameter for selected species

Mean Height Comparisons

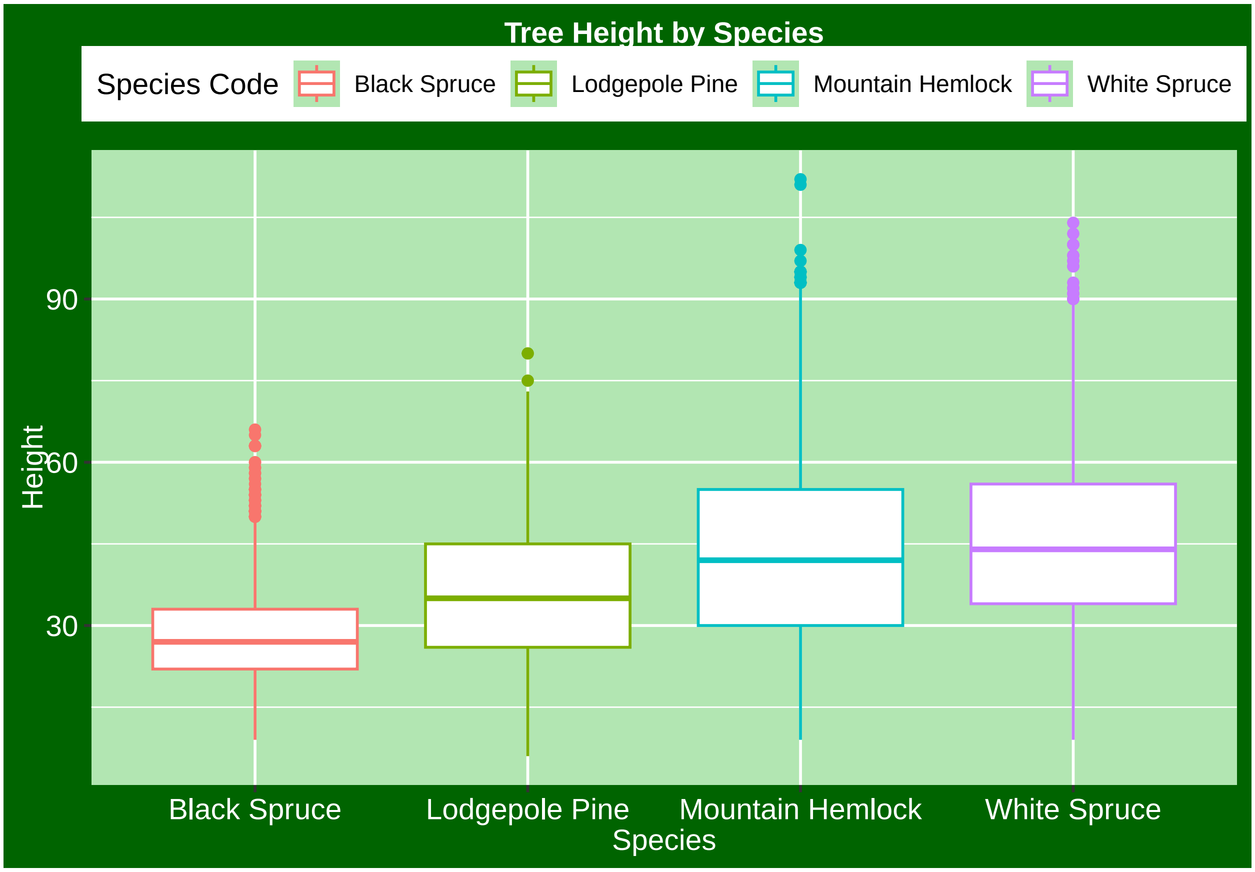

Figure 3: This figure displays a bloxplot showing the differences in mean and distribution between height for select species

The boxplot created here was key in visualizing the distribution of data for each species selected. Also gave an idea of how different each mean height of the different species were. Allowing clarification in this overloaded dataset, before moving on into deeper analysis.

Question 3: How does height vary with plot?

Correlation between Height and Diameter taking Species of Interest and Plot ID into Account

Unfortunately, there are too many unique plots to visualize this linear regression, but the results indicate that, in general, plot and species have a significant influence on the height of the tree (with the p-value for most plots being <0.01). These results will be confirmed with an ANOVA in the following analysis section. This linear regression was run but results are to long to show in the document. The results showed most of the plots significantly influenced height so we included it in the anova. The code for this output is shown at the very end of the assignment if needed.

Analysis

The following steps are the bulk of our analysis. These ANOVA tests, Tukey HSD tests, and visualizations allowed us to determine whether or not these various factors influenced tree height.

Questions 1,2, and 3: How does height vary by DBH, species, and plot?

ANOVA Test for Height Infuenced by Diameter and Species

```
##                Df  Sum Sq Mean Sq F value Pr(>F)    
## DIA             1 1351652 1351652 17894.8 <2e-16 ***
## Species         3  216899   72300   957.2 <2e-16 ***
## Residuals    8534  644599      76                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 82 observations deleted due to missingness
```

ANOVA Test for Height Influenced by Diameter, Species, and Plot

```
##                Df  Sum Sq Mean Sq   F value Pr(>F)    
## DIA             1 1351652 1351652 47462.200 <2e-16 ***
## Species         3  216899   72300  2538.752 <2e-16 ***
## PLOT         1934  456641     236     8.291 <2e-16 ***
## Residuals    6600  187958      28                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 82 observations deleted due to missingness
```

The ANOVA results suggest that diameter has the highest influence on tree height (DF= 1 and p-value <0.01, F-value=17894.8), as expected. Additionally, it suggests that both species (DF=3, p < 0.01, F-value=957.2) and plot number (DF=8534, p<0.01, F-value=8.291) have a significant influence on tree height. The influence of plot is smaller than that of diameter and tree species. These results will be explored further in the following sections.

Question 2: How does height vary by species?

ANOVA Test to Compare Mean Height Differences Between Species

In order to properly understand the statistical differences present in mean species height, an ANOVA test will be conducted between each species in pairs.

```
##                Df Sum Sq Mean Sq F value Pr(>F)    
## Species         1 363941  363941    2077 <2e-16 ***
## Residuals    4951 867673     175                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)    
## Species         1  59934   59934   257.6 <2e-16 ***
## Residuals    3669 853755     233                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Species        1  33524   33524   129.7 <2e-16 ***
## Residuals   3666 947667     259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Species        1 270470  270470    1392 <2e-16 ***
## Residuals   4948 961585     194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Species        1  38618   38618   361.7 <2e-16 ***
## Residuals   3052 325847     107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##               Df  Sum Sq Mean Sq F value   Pr(>F)
## Species        1    7881    7881   29.45 5.99e-08 ***
## Residuals   5565 1489494     268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Each ANOVA pairing between species is significantly different, having a p value below 0.05. This means that the null hyptohesis is rejected, and the alternative hypothesis is accepted: tree height does vary significantly between species.

Because the anova test shows that all results are significant a pairwise comparison between species can be conducted using a Tukey Test HSD to show the confidence interval, p-value and difference in each species.

Tukey Test HSD of Tree Species Mean Height Difference

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = HT ~ Species, data = w.spruce.l.pine)
##
## $Species
##                                  diff      lwr      upr p adj
## White Spruce-Lodgepole Pine 9.442757 8.289179 10.59634     0
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = HT ~ Species, data = l.pine.m.hemlock)
##
## $Species
##                                      diff      lwr      upr p adj
## Mountain Hemlock-Lodgepole Pine 7.063093 5.847069 8.279118     0
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
##
## Fit: aov(formula = HT ~ Species, data = b.spruce.m.hemlock)
##
## $Species
##                                    diff      lwr      upr p adj
## Mountain Hemlock-Black Spruce 14.89889 14.11595 15.68182     0
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = HT ~ Species, data = b.spruce.l.pine)
##
## $Species
##                                  diff      lwr      upr p adj
## Lodgepole Pine-Black Spruce 7.835792 7.027957 8.643627     0
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = HT ~ Species, data = b.w.spruces)
##
## $Species
##                                diff      lwr      upr p adj
## White Spruce-Black Spruce 17.27855 16.53523 18.02187     0
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = HT ~ Species, data = w.spruce.m.hemlock)
##
## $Species
##                                    diff     lwr      upr p adj
## White Spruce-Mountain Hemlock 2.379664 1.51996 3.239367 1e-07
```

```
##                                    diff       lwr       upr        p adj
## White Spruce-Lodgepole Pine      9.442757  8.289179 10.596335 8.369286e-09
## Mountain Hemlock-Lodgepole Pine  7.063093  5.847069  8.279118 8.252223e-09
## Mountain Hemlock-Black Spruce   14.898886 14.115947 15.681825 3.501914e-09
## Lodgepole Pine-Black Spruce      7.835792  7.027957  8.643627 0.000000e+00
## White Spruce-Black Spruce       17.278550 16.535226 18.021873 2.937114e-09
## White Spruce-Mountain Hemlock    2.379664  1.519960  3.239367 5.994149e-08
##                                                      Comparison
## White Spruce-Lodgepole Pine         White Spruce vs Lodgepole Pine
## Mountain Hemlock-Lodgepole Pine Lodgepole Pine vs Mountain Hemlock
## Mountain Hemlock-Black Spruce     Black Spruce vs Mountain Hemlock
## Lodgepole Pine-Black Spruce         Black Spruce vs Lodgepole Pine
## White Spruce-Black Spruce             White Spruce vs Black Spruce
## White Spruce-Mountain Hemlock     White Spruce vs Mountain Hemlock
```

According to the Tukey HSD Test, the largest difference in tree height means is between black spruce and white spruce (17.28 feet), followed by the difference of Mountain Hemlock and Black Spruce (14.89 feet). The trees with the least mean height difference are White Spruce and Mountain Hemlock (2.38 feet). The confidence intervals between true differences in height ranged throughout species with the biggest confidence interval being Mountain Hemlock versus Lodgepole Pine with a lower bound of 5.84 feet and a higher bound of 10.59 feet. The smallest confidence interval was found between Mountain Hemlock and Black Spruce and Lodgepole Pine and Black Spruce (14.11 to 15.69 feet, and 7.03 and 8.64 feet, respectively). Lastly, all p-values have stayed statistically significant. This means we can reject the null hypothesis and accept the alternative hypothesis: tree species in Alaska do vary by height.

Questoin 3: How does height vary by plot?

Visualizing Interactions with Plot Differences

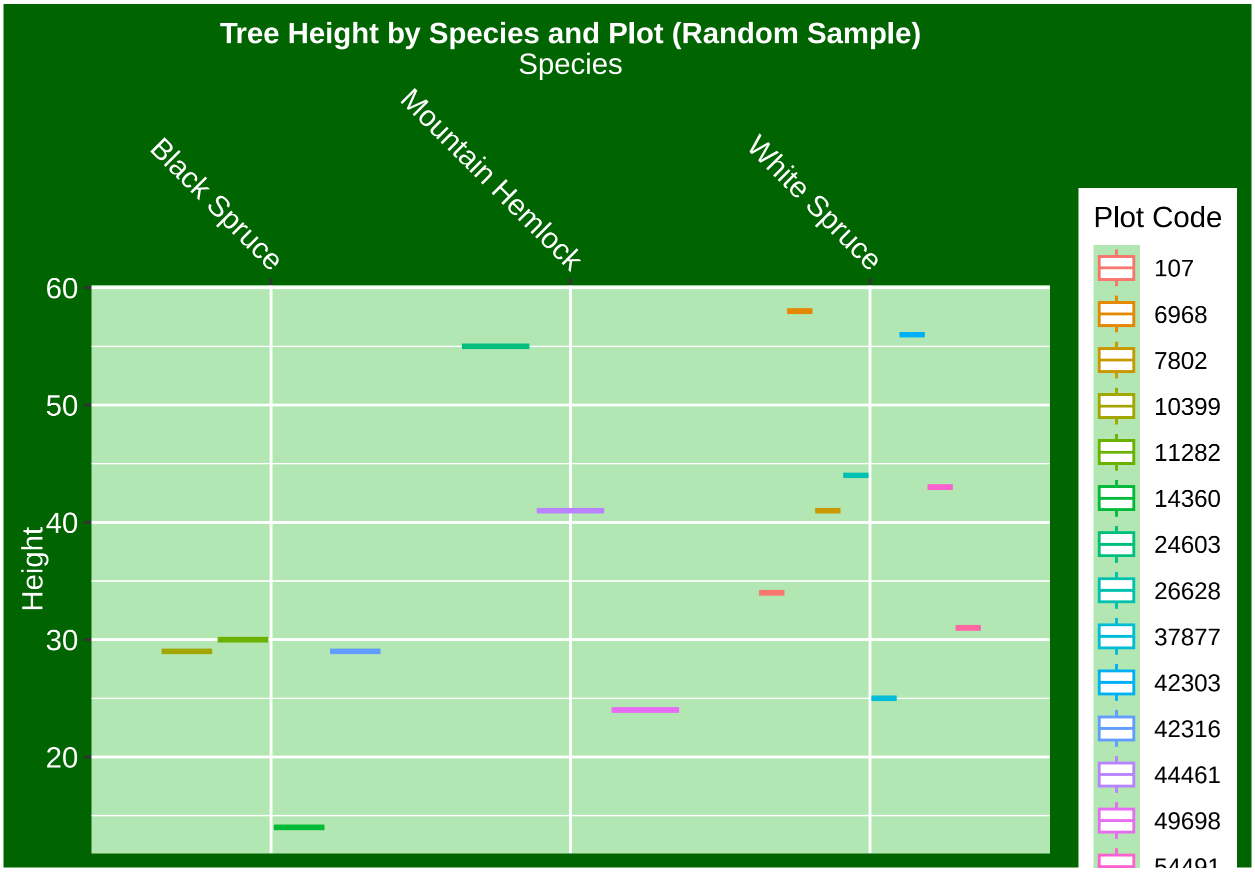

Figure 4: This figure displays a random sample of Plots from our dataset. As shown in the figure, the average heights vary both by species and by plot, demonstrating that plot is a significant influencing factor on tree height along with species and diameter

This data visualization emphasizes the results from the ANOVA tests, showing that plot is also a significant influencing factor on tree height. There were too many unique plots to visualize the results from the linear regression or show all plots in this figure, but even this random sample reemphasizes the findings from our ANOVA test.

Summary and Conclusions

Through our analysis, we came to the conclusion that the trees in our Alaskan dataset differ significantly by height, allowing us to reject our null hypothesis. In specifics, diameter had the most significant effect on height, with plotID having the least significant effect, and species in between.

Through our interest in the species diversity in Alaskan forests, questioned on how tree height varies, with an emphasis on how diameter, species, and plot ID influence this height variation. After wrangling the data for four specific species to use in our analysis, we created correlations to find exploratory trends between the variables. An example of one of the trends we found was the the the height of black spruce trees is less affected by diameter than other species.

To determine significance in height between species, we ran an ANOVA for each unique pair of species, a total of six ANOVAs. Because each pair displayed a significant p-value, we ran Tukey HSD tests to determine not only which species pairs were the most significantly different, but also the magnitude of their differences in average height. The most significant difference was white spruce and black spruce, with an average height difference of 17.28 ft. The least significant difference was between white spruce and mountain hemlock (2.38 ft). These significant differences in tree height confirmed the rejection of our null hypothesis. Overall, we learned that diameter at breast height, species, and plotID all had significant impacts on tree height, though to a varying degree, allowing us to reject the null hypothesis.

References

U.S. Forest Service. (2020). Forest Inventory and Analysis Database (FIADB) user guides: Volume database description (version 9.2). U.S. Department of Agriculture. Retrieved December 11, 2024, from https://www.fia.fs.fed.us/library/database-descriptions/

U.S. Forest Service. (n.d.). Interior Alaska Inventory - Pacific Northwest Forest Inventory and Analysis (PNW-FIA). U.S. Department of Agriculture. Retrieved December 11, 2024, from https://research.fs.usda.gov/pnw/projects/pnwfiainterioralaskainventory

Appendix

```r
#Loading Packages
knitr::opts_chunk$set(echo = TRUE)
library(dplyr)
library(ggplot2)
library(tidyverse)

# Set your working directory
getwd()
setwd("/home/guest/EDEProj_Arunkumar_Power_Rotbart_Valkenberg")

# Load your datasets
Ak_data <-
  read.csv(
    '~/EDEProj_Arunkumar_Power_Rotbart_Valkenberg/Data_Overall/Raw/AK_SITETREE.csv')

#Filtering for wanted columns
Ak_data.wrangled <- Ak_data %>%
  select(PLOT, SPCD, DIA, HT)

#Filtering for Species
Species.wanted <- Ak_data.wrangled %>%
  filter(SPCD %in% c(94, 95, 108, 264))
Species.wanted <- Species.wanted %>%
  mutate(Species = case_when(
    SPCD == 94 ~ "White Spruce",
    SPCD == 95 ~ "Black Spruce",
    SPCD == 108 ~ "Lodgepole Pine",
    SPCD == 264 ~ "Mountain Hemlock"
  ))
Species.wanted$Species <- as.factor(Species.wanted$Species)
Species.wanted$PLOT <- as.factor(Species.wanted$PLOT)

write.csv(Species.wanted,
file =
  '~/EDEProj_Arunkumar_Power_Rotbart_Valkenberg/Data_Overall/Processed/AK_SITETREE_SpeciesWanted.csv',
row.names = FALSE)
# Set your ggplot theme

our_theme <- theme(
  axis.text = element_text(color = "white"),
```

```r
  legend.position = "top",
  plot.background = element_rect(fill = "#006400",color = NA),
  panel.background = element_rect(fill = "#b2e6b2", color = NA),
  plot.title = element_text(face = "bold", color="white", hjust = 0.5),
  plot.subtitle = element_text(face = "bold", color="white", hjust = 0.5),
  axis.title.x = element_text(color="white"),
  axis.title.y = element_text(angle=90, color="white")
)
theme_set(our_theme)
#Linear Regression to see correlation between HT and DIA for all Species in Data
All.species.AK <- lm(HT ~ DIA, data = Ak_data.wrangled)
summary(All.species.AK)
ggplot(Ak_data.wrangled, aes(x = DIA, y = HT)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Tree Height vs Diameter in Alaska",
       x = "Diameter",
       y = "Height")
#Linear Regression to see correlation between HT and DIA taking into account
#four species.
species.model <- lm(HT ~ DIA + Species, data = Species.wanted)
summary(species.model)

ggplot(Species.wanted, aes(x = DIA, y = HT, color = Species)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Tree Height vs Diameter by Species",
       x = "Diameter",
       y = "Height",
       color = "Species Code")

#Boxplot of Tree Height by Species
ggplot(Species.wanted, aes(x = Species, y = HT, color = Species)) +
  geom_boxplot()+
  labs(title = "Tree Height by Species",
       x = "Species",
       y = "Height",
       color = "Species Code")
#Linear regression to see correlation between HT and DIA taking into account
#four species and plot ID
PlotandSpecies.model <- lm(HT ~ DIA + Species + PLOT, data = Species.wanted)
summary(PlotandSpecies.model)


#ANOVA with species interaction
species.model.anova <- aov(HT ~ DIA + Species,
                           data = Species.wanted)
summary(species.model.anova)



#ANOVA with species and plot interaction
PlotandSpecies.model.anova <- aov(HT ~ DIA + Species + PLOT,
```

```r
                                  data = Species.wanted)
summary(PlotandSpecies.model.anova)

#Mean Height of Black Spruce compared to Mean Height of White Spruce
b.w.spruces <- Species.wanted %>%
  filter(SPCD %in% c(94, 95))
BlackvWhiteSpruce <- aov(HT~Species, data = b.w.spruces)

summary.aov(BlackvWhiteSpruce)

#Mean Height of White Spruce compared to Mean Height of Lodgepole Pine
w.spruce.l.pine <- Species.wanted %>%
  filter(SPCD %in% c(94, 108))
WhiteSprucevLodgePine <- aov(HT~Species, data = w.spruce.l.pine)

summary.aov(WhiteSprucevLodgePine)

#Mean Height of Lodgepole Pine compared to Mean Height of Mountain Hemlock
l.pine.m.hemlock <- Species.wanted %>%
  filter(SPCD %in% c(108, 264))
LodgePinevMountHemlock <- aov(HT~Species, data = l.pine.m.hemlock)

summary.aov(LodgePinevMountHemlock)

#Mean Height of Black Spruce compared to Mean Height of Mountain Hemlock
b.spruce.m.hemlock <- Species.wanted %>%
  filter(SPCD %in% c(95, 264))
BlackSprucevMountHemlock <- aov(HT~Species, data = b.spruce.m.hemlock)

summary.aov(BlackSprucevMountHemlock)

#Mean Height of Black Spruce compared to Mean Height of Lodgepole Pine
b.spruce.l.pine <- Species.wanted %>%
  filter(SPCD %in% c(95, 108))
BlackSprucevLodgePine <- aov(HT~Species, data = b.spruce.l.pine)

summary.aov(BlackSprucevLodgePine)

#Mean Height of White Spruce compared to Mean Height of Mountain Hemlock
w.spruce.m.hemlock <- Species.wanted %>%
  filter(SPCD %in% c(94, 264))
WhiteSprucevMountHemlock <- aov(HT~Species, data = w.spruce.m.hemlock)

summary.aov(WhiteSprucevMountHemlock)
#Conduct Tukey test on each species pair
TukeyResults1 <- TukeyHSD(WhiteSprucevLodgePine)
TukeyResults2 <- TukeyHSD(LodgePinevMountHemlock)
TukeyResults3 <- TukeyHSD(BlackSprucevMountHemlock)
TukeyResults4 <- TukeyHSD(BlackSprucevLodgePine)
TukeyResults5 <- TukeyHSD(BlackvWhiteSpruce)
TukeyResults6 <- TukeyHSD(WhiteSprucevMountHemlock)

#Tukey result objects
```

```r
TukeyResults1
TukeyResults2
TukeyResults3
TukeyResults4
TukeyResults5
TukeyResults6

#Tukey data frame to bind and compare Tukey results
TukeyResults1_df <- as.data.frame(TukeyResults1$Species) %>%
  mutate(Comparison = "White Spruce vs Lodgepole Pine")

TukeyResults2_df <- as.data.frame(TukeyResults2$Species) %>%
  mutate(Comparison = "Lodgepole Pine vs Mountain Hemlock")

TukeyResults3_df <- as.data.frame(TukeyResults3$Species) %>%
  mutate(Comparison = "Black Spruce vs Mountain Hemlock")

TukeyResults4_df <- as.data.frame(TukeyResults4$Species) %>%
  mutate(Comparison = "Black Spruce vs Lodgepole Pine")

TukeyResults5_df <- as.data.frame(TukeyResults5$Species) %>%
  mutate(Comparison = "White Spruce vs Black Spruce")

TukeyResults6_df <- as.data.frame(TukeyResults6$Species) %>%
  mutate(Comparison = "White Spruce vs Mountain Hemlock")

all_TukeyResults <- bind_rows(TukeyResults1_df,TukeyResults2_df, TukeyResults3_df, TukeyResults4_df, Tuk

#Tukey dataframe results for comparisons between all species
print(all_TukeyResults)

random_plot_subset <- Species.wanted[sample(nrow(Species.wanted), size = 15,
                                            replace = FALSE), ]

#Plot showing relationship between HT, PLOT, and Species (random sample)
ggplot(random_plot_subset, aes(x = Species, y = HT, color = PLOT)) +
  geom_boxplot() +
  labs(title = "Tree Height by Species and Plot (Random Sample)",
       x = "Species",
       y = "Height",
       color = "Plot Code") +
  theme(axis.text.x = element_text(angle = -45, hjust = 1)) +
  theme(legend.position = "right")+
  scale_x_discrete(position = "top")
```