Lab 2: Minnesota Tree Growth

AUTHOR
Hanna Velicer

PUBLISHED
February 21, 2025

Question 1

Read in the Minnesota tree growth dataset. Use `glimpse` to understand the structure and names of the dataset. Decribe the structure and what you see in the dataset?

```
library(dplyr)
tree_dat <- read.csv("/Users/hannavelicer/Desktop/ESS330/lab2-tree-growth/data/tree_dat.c
glimpse("tree_dat")
```

```
 chr "tree_dat"
```

Answer: This dataset is a 131,386 x 8 table of tree growth analysis for 15 different species. The data ranges from 1897 to 2007. The columns are treeID, standID, stand, year, species, age, inc, and rad_ib.

Question 2

How many records have been made in stand 1?

```
tree_dat %>%
  filter(standID == 1) %>%
  tally()
```

```
    n
1 979
```

Answer: There have been 979 records made in stand 1.

Question 3

How many records of the Abies balsamea and Pinus strobus species have been made?

```
tree_dat %>%
  filter(species %in% c("ABBA","PIST")) %>%
  count(species)
```

```
  species     n
1    ABBA 13033
2    PIST  4188
```

Answer: There are 13,033 records of the Abies balsamea and 4,188 records of the Pinus strobus.

Question 4

How many trees are older then 200 years old in the last year of the dataset?

```
last_year <- max(tree_dat$year, na.rm = TRUE)
tree_dat %>%
  filter(year == last_year, age > 200) %>%
  tally()
```

```
  n
1 7
```

Answer: There are 7 trees older than 200 years old in the last year of the dataset.

Question 5

What is the oldest tree in the dataset found using slice_max?

```
tree_dat %>%
  slice_max(order_by = age, n = 1)
```

```
  treeID standID stand year species age  inc rad_ib
1     24       2    A2 2007    PIRE 269 0.37 308.84
```

Answer: The oldest tree in the dataset is 269 years old.

Question 6

Find the oldest 5 trees recorded in 2001. Use the help docs to understand optional parameters.

```
tree_dat %>%
  filter(year == 2001) %>%
  slice_max(order_by = age, n = 5)
```

```
  treeID standID stand year species age   inc  rad_ib
1     24       2    A2 2001    PIRE 263 0.210 306.880
2     25       2    A2 2001    PIRE 259 0.280 156.210
3   1595      24    F1 2001    FRNI 212 0.579 156.267
4   1598      24    F1 2001    FRNI 206 0.394 130.251
5   1712      26    F3 2001    FRNI 206 0.168 154.354
```

Answer: The oldest 5 trees in 2001 were three Fraxinus nigra (*212, 206, 206*) and two Pinus resinosa (*2633, 259*).

Question 7

Using slice_sample, how many trees are in a 30% sample of those recorded in 2002?

```
tree_dat %>%
  filter(year == 2002) %>%
  slice_sample(prop = 0.3)
```

Answer: There are 687 trees in a 30% sample of those recorded in 2002. (*for the sake of length, I chose to hide the coding result of this question as it is 687 rows and a pain to scroll through!*)

Question 8

Filter all trees in stand 5 in 2007. Sort this subset by descending radius at breast height (rad_ib) and use slice_head() to get the top three trees. Report the tree IDs.

```
tree_dat %>%
  filter(standID == 5, year == 2007) %>%
  arrange(desc(rad_ib)) %>%
  slice_head(n = 3)
```

```
  treeID standID stand year species age   inc   rad_ib
1    128       5    A6 2007    PIST  82 0.885 238.8850
2    157       5    A6 2007    PIRE  85 0.900 217.8700
3    135       5    A6 2007    PIMA  84 0.110 210.1874
```

Answer: The top 3 trees are #128, #157, and #135.

Question 9

Reduce your full `data.frame` to [treeID, stand, year, and radius at breast height]. Filter to only those in stand 3 with records from 2007, and use slice_min to pull the smallest three trees measured that year.

```
tree_dat %>%
  select(treeID, standID, year, rad_ib) %>%
  filter(standID == 3, year == 2007) %>%
  slice_min(rad_ib, n = 3)
```

```
  treeID standID year rad_ib
1     50       3 2007 47.396
2     56       3 2007 48.440
3     36       3 2007 54.925
```

Answer: The smallest 3 trees are #50, #56, and #36.

Question 10

Use select to remove the stand column. Use `glimspe` to show the dataset.

```
tree_dat %>%
  select(-stand) %>%
  glimpse()
```

```
Rows: 131,386
Columns: 7
$ treeID  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
$ standID <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
$ year    <int> 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 19…
$ species <chr> "ABBA", "ABBA", "ABBA", "ABBA", "ABBA", "ABBA", "ABBA", "ABBA"…
$ age     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,…
$ inc     <dbl> 0.930, 0.950, 0.985, 0.985, 0.715, 0.840, 0.685, 0.940, 1.165,…
$ rad_ib  <dbl> 10.78145, 11.73145, 12.71645, 13.70145, 14.41645, 15.25645, 15…
```

Question 11

Look at the help document for dplyr::select and examine the "Overview of selection features". Identify an option (there are multiple) that would help select all columns with the string "ID" in the name. Using glimpse to view the remaining dataset.

```
?dplyr::select
tree_dat %>%
  select(contains("ID")) %>%
  glimpse()
```

```
Rows: 131,386
Columns: 2
$ treeID  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
$ standID <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
```

Question 12

Find a selection pattern that captures all columns with either 'ID' or 'stand' in the name. Use glimpse to verify the selection.

```
tree_dat %>%
  select(matches("ID|stand")) %>%
  glimpse()
```

```
Rows: 131,386
Columns: 3
```

```
$ treeID  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
$ standID <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
$ stand   <chr> "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A…
```

Question 13

Looking back at the data dictionary, rename rad_inc and inc to include `_[unit]` in the name. Unlike earlier options, be sure that this renaming is permanent, and stays with your data.frame (e.g. `<-`). Use glimpse to view your new data.frame.

```
tree_dat <- tree_dat %>%
  rename(rad_ib_mm = rad_ib, inc_mm = inc) %>%
  glimpse()
```

```
Rows: 131,386
Columns: 8
$ treeID    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
$ standID   <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
$ stand     <chr> "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", "A1", …
$ year      <int> 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, …
$ species   <chr> "ABBA", "ABBA", "ABBA", "ABBA", "ABBA", "ABBA", "ABBA", "ABB…
$ age       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 1…
$ inc_mm    <dbl> 0.930, 0.950, 0.985, 0.985, 0.715, 0.840, 0.685, 0.940, 1.16…
$ rad_ib_mm <dbl> 10.78145, 11.73145, 12.71645, 13.70145, 14.41645, 15.25645, …
```

Question 14

A key measurement in forestry in “basal area column”. The metric is computed with the formula: BA(m2) = 0.00007854•DBH^2, where DBH is the diameter at breast height (cm). Use mutate to compute DBH in centimeters, and BA in m2 (HINT: Make sure rad_ib is in cm prior to computing the diameter!). What is the mean BA_m2 of the the species `POTR` in 2007?

```
tree_dat <- tree_dat %>%
  mutate(rad_ib_cm = rad_ib_mm/10) %>%
  mutate(DBH_cm = rad_ib_cm * 2, BA_m2 = 0.00007854 * (DBH_cm^2))
tree_dat %>%
  filter(species == "POTR", year == 2007) %>%
  summarize (mean_BA_m2 = mean(BA_m2, na.rm = TRUE)) %>%
  print()
```

```
  mean_BA_m2
1 0.03696619
```

Answer: The mean BA_m2 of the `POTR` species in 2007 was 0.03696619 m2.

Question 15

Lets say for the sake of our study, trees are not established until they are 5 years of age. Use `if_else` to add a boolean column to our dataset called `established` that is TRUE if the age is greater then 5 and FALSE if less then or equal to five. Once added, use count (see ?count) to determine how many records are from established trees?

```
tree_dat %>%
  mutate(established = if_else(age > 5, TRUE, FALSE)) %>%
  count(established)
```

```
  established      n
1       FALSE   8883
2        TRUE 122503
```

Answer: There are 122,503 records from "established" trees.

Question 16

Use `mutate` and `case_when` to add a new column to you data.frame that classifies each tree into the proper DBH_class. Once done, limit your dataset to the year 2007 and report the number of each class with count.

```
tree_dat %>%
  mutate(DBH_class = case_when(
    DBH_cm < 2.5 ~ "Seedling",
    DBH_cm >= 2.5 & DBH_cm < 10 ~ "Sapling",
    DBH_cm >= 10 & DBH_cm < 30 ~ "Pole",
    DBH_cm >= 30 ~ "Sawlog",
  )) %>%
  filter(year == 2007) %>%
  count(DBH_class) %>%
  print()
```

```
  DBH_class    n
1      Pole 1963
2   Sapling  252
3    Sawlog   76
```

Answer: The "Sapling" class contains 252 trees, the "Pole" class contains 1,963 trees, the "Sawlog" class contains 76 trees, and the "Seedling" class contains 0 trees.

Question 17

Compute the mean DBH (in cm) and standard deviation of DBH (in cm) for all trees in 2007. Explain the values you found and their statistical meaning.

```
tree_dat %>%
  filter(year == 2007) %>%
  summarize(mean_DBH_cm = mean(DBH_cm, na.rm = TRUE), sd_DBH_cm = sd(DBH_cm, na.rm = TRUE
  print()
```

```
  mean_DBH_cm sd_DBH_cm
1    16.09351  6.138643
```

Answer: The mean DBH is 16.09351 cm, which is the average tree diameter at breast height in 2007. The standard deviation of DBH is 6.138643 cm, which is the spread of tree sizes around the mean. This number suggests that there was a large variation in tree sizes.

Question 18

Compute the per species mean tree age using only those ages recorded in 2003. Identify the three species with the oldest mean age.

```
tree_dat %>%
  filter(year == 2003) %>%
  group_by(species) %>%
  summarize(mean_age = mean(age, na.rm = TRUE)) %>%
  arrange(desc(mean_age)) %>%
  slice_head(n = 3) %>%
  print()
```

```
# A tibble: 3 × 2
  species mean_age
  <chr>      <dbl>
1 THOC       127. 
2 FRNI        83.1
3 PIST        73.3
```

Answer: The three species with the oldest mean age are Thuja occidentalis (*127*), Fraxinus nigra (*83*), and Pinus strobus (*73*).

Question 19

In a single summarize call, find the number of unique years with records in the data set along with the first and last year recorded?

```
tree_dat %>%
  summarize(unique_years = n_distinct(year), first_year = min(year, na.rm = TRUE), last_y
  print()
```

```
  unique_years first_year last_year
1          111       1897      2007
```

Answer: There are 111 unique years in the dataset, with 1897 being the first year with recorded data and 2007 being the last.

Question 20

Determine the stands with the largest number of unique years recorded. Report all stands with largest (or tied with the largest) temporal record.

```
stand_years <- tree_dat %>%
  group_by(stand) %>%
  summarize(unique_years = n_distinct(year)) %>%
  arrange(desc(unique_years))
max_years <- max(stand_years$unique_years)
top_stands <- stand_years %>%
  filter(unique_years == max_years) %>%
  print()
```

```
# A tibble: 5 × 2
  stand unique_years
  <chr>        <int>
1 A1             111
2 D1             111
3 D2             111
4 D3             111
5 F1             111
```

Answer: There are 5 stands tied for the largest amount of unique years (*111*): A1, D1, D2, D3, and F1.

Final Question

We are interested in the annual DBH growth rate of each species through time, but we only want to include trees with at least a 10 year growth record. To identify this, we need to identify the per year growth made by each tree, their total growth record, and then average that, and compute the standard deviation, across the species. Use a combination of dplyr verbs to compute these values and report the 3 species with the fastest growth, and the 3 species with the slowest growth. (** You will need to use either `lag()` or `diff()` in your computation. You can learn more about each in the Help pages) Lastly, find and include an image of the fastest growing species. Add the image to your images directory.

```
tree_dat %>%
  group_by(treeID) %>%
  arrange(year) %>%
  mutate(growth = DBH_cm - lag(DBH_cm),
         total_years = n()) %>%
  filter(!is.na(growth), total_years > 10) %>%
```

```
  ungroup() %>%
  group_by(species) %>%
  summarize(mean_growth = mean(growth)) %>%
  slice_max(mean_growth, n = 3) %>%
  print() %>%
  slice_min(mean_growth, n = 3) %>%
  print()
```

```
# A tibble: 3 × 2
  species mean_growth
  <chr>         <dbl>
1 PIRE          0.358
2 POTR          0.331
3 PIBA          0.326
# A tibble: 3 × 2
  species mean_growth
  <chr>         <dbl>
1 PIBA          0.326
2 POTR          0.331
3 PIRE          0.358
```

The 3 species with the fastest growth are Pinus resinosa (*0.357, overall fastest*), Populus tremuloides (*0.330*), and Pinus banksiana (*0.325*). The 3 species with the slowest growth are Larix laricina (*0.149*), Thuja occidentalis (*0.152*), and Quercus rubra (*0.167*).

Pinus resinosa

This image depicts Pinus resinosa, commonly known as red pine, which is the tree with the overall fastest growth.