

MEP - Forest Biomass dynamics

Session 1 - Practical worksheet

19th October 2022

1 Introduction

You will need to use R and a spreadsheet program like Microsoft Excel to complete this practical session.

Please speak up if you don't have access to the relevant software.

The teal coloured text, e.g. `EXAMPLE example`, refers to R code, variable names, or file names.

The boxes contain questions or tasks for you to complete on your own. Record your thoughts and answers to each question in a notebook or a word document.

In this practical you will:

- practice manipulating forest tree stem census data in R.
- estimate tree biomass using an allometric equation.
- calculate growth, recruitment, and mortality rates.
- compare sites across environmental gradients to determine drivers of biomass dynamics.
- calculate species and size-class specific growth rates.
- predict the biomass dynamics of an unmeasured site.

2 Datasets

You can download all the data needed for this practical from the LEARN page for this session. Save each of the files to a folder on your computer. There is also an R script (`example_script.R`) which contains examples of all the code you should need in this practical. Try not to rely on this too much, only have a peek if you are really stuck.

`stems.csv` contains a table of tree stem measurements from 10 forest sites across the Brazilian Amazon. The data are synthetic, but approximate real data in terms of species composition, population dynamics, and number of stems. Each site consists of a 1 ha (100x100 m) square plot. Each site received 10 censuses, one every five years, where all tree stems >5 cm diameter were measured. Each row is a stem measurement at a particular time point. When a tree dies, it is given diameter and height measurements of 0 (zero) for that census and all future censuses. Here is a description of the columns in the table:

- `site_id` - ID of the site
- `species_name` - species name
- `diam` - stem diameter (cm)
- `height` - tree height (m)
- `census_date` - year of the census
- `stem_id` - ID of the stem

`sites.csv` contains a table of site information for each of the sites in `stems.csv`. The `site_id` column links to the `site_id` column in `stems.csv`. The `lon` and `lat` columns contain the longitude and latitude coordinates of the site, respectively. The remaining columns (beginning with `bio`) contain average climate data for each site, which was extracted from the “BioClim” dataset (<https://www.worldclim.com/bioclim>). The BioClim data are calculated as the mean of annual values for the years 1970-2000, and have been aggregated to 0.083×0.083 degree cells, which is approximately 8 km² at the equator. Here is a description of each of those columns:

- `bio1` = Annual Mean Temperature
- `bio2` = Mean Diurnal Range (Mean of monthly (max. temp–min. temp))
- `bio3` = Isothermality (`bio2/bio7`) ($\times 100$)

- `bio4` = Temperature Seasonality (standard deviation $\times 100$)
- `bio5` = Max Temperature of Warmest Month
- `bio6` = Min Temperature of Coldest Month
- `bio7` = Temperature Annual Range (`bio5-bio6`)
- `bio8` = Mean Temperature of Wettest Quarter (3 months)
- `bio9` = Mean Temperature of Driest Quarter (3 months)
- `bio10` = Mean Temperature of Warmest Quarter (3 months)
- `bio11` = Mean Temperature of Coldest Quarter (3 months)
- `bio12` = Total Annual Precipitation
- `bio13` = Precipitation of Wettest Month
- `bio14` = Precipitation of Driest Month
- `bio15` = Precipitation Seasonality (Coefficient of Variation)
- `bio16` = Precipitation of Wettest Quarter (3 months)
- `bio17` = Precipitation of Driest Quarter (3 months)
- `bio18` = Precipitation of Warmest Quarter (3 months)
- `bio19` = Precipitation of Coldest Quarter (3 months)

3 Loading data

First, open a new script in R (I recommend using RStudio) and set the working directory (e.g. `setwd()`) to the location of `stems.csv` and `sites.csv`.

Then, load the data into R:

```
stems <- read.csv("./stems.csv")
sites <- read.csv("./sites.csv")
```

Have a look at the data to understand its structure. Confirm that all the columns listed above are present in the dataframes:

```
str(stems)
head(stems)

str(sites)
head(sites)
```

Make some quick histograms of the data to understand the distributions of each variable, for example:

```
hist(stems$diam)
```

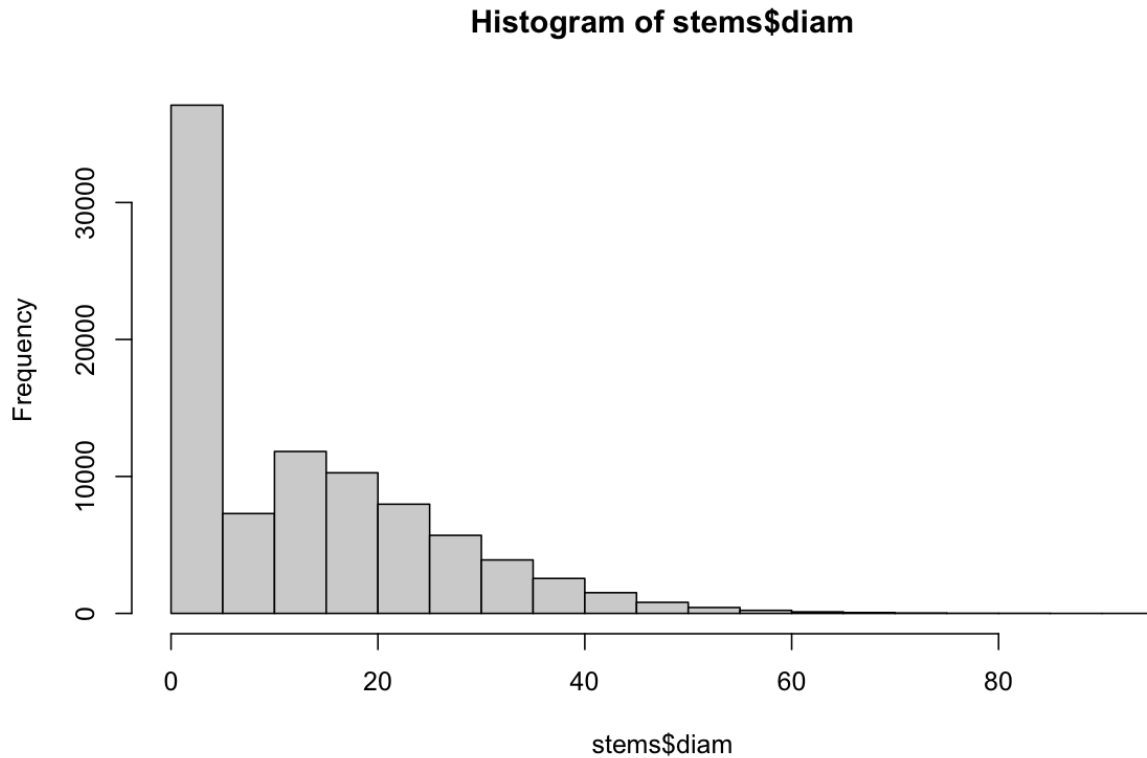


Figure 1: Histogram of stem diameters.

4 Calculating stem-level biomass using an allometric equation

Our ultimate goal is to describe the biomass dynamics of the region, and understand variation in biomass dynamics among sites. Firstly then, we need to estimate the biomass contained in each tree stem using the field measurements in `stems`.

We can get wood density (g cm^{-3}) for each species in our dataset using an online database. Load the `{BIOMASS}` package:

```
library(BIOMASS)
```

If the package doesn't load, you may need to install it first, using `install.packages("BIOMASS")`.

Look at the help file for the `getWoodDensity()` function:

```
?getWoodDensity
```

The function takes arguments of `genus`, `species`, `stand` (site), and other optional arguments.

We need to split the species names in `stems$species_name` to get a separate column for `genus` and `species`. The code below takes everything before or after the “space” character (`\\s`) and replaces it with nothing (`""`). E.g. “Burkea africana” becomes genus: “Burkea”, species: “africana”:

```
stems$genus <- gsub("\\s.*", "", stems$species_name)
stems$species <- gsub(".*\\s", "", stems$species_name)
```

Now we can get the wood density values for each individual:

```
wd <- getWoodDensity(stems$genus, stems$species, stems$site_id,
  region = "SouthAmericaTrop")
```

`wd` contains a dataframe with the mean wood density (`wd$meanWD`) for each individual in `stems`. So we can add `wd` to `stems`:

```
stems$wd <- wd$meanWD
```

Next we can use the pan-tropical allometric equation from Chave et al. (2014) to estimate the biomass (AGB, above-ground biomass) of each stem:

$$B = 0.0673 \times (\rho \times H \times D^2)^{0.976} / 1000$$

```
stems$agb <- (0.0673 * (stems$wd * stems$height * stems$diam^2)^0.976)/1000
```

There are many sources of uncertainty in this empirical model used to estimate biomass. List as many as you can.

5 Visualising changes in demography over time

Next, let's have a look at how biomass and number of stems changes within each site over time. To summarise the data we will use the `{dplyr}` and `{tidyr}` packages, which provide many handy functions for manipulating tabular data.

```
library(dplyr)
library(tidyr)
```

The code below counts the number of living stems for each census within each site.

```
stems_ab_summ <- stems %>%
  group_by(site_id, census_date) %>%
  summarise(n_alive = sum(diam != 0))
```

There should be 100 rows in `stems_ab_summ`, as we have 10 sites, each with 10 censuses.

Now we can create a plot of number of stems over time for each site:

```
ggplot(data = stems_ab_summ, # dataframe
  aes(x = census_date, y = n_alive, colour = site_id)) + # Axes and groupings
  geom_point() + # Adds points
  geom_path() # Adds lines between points
```

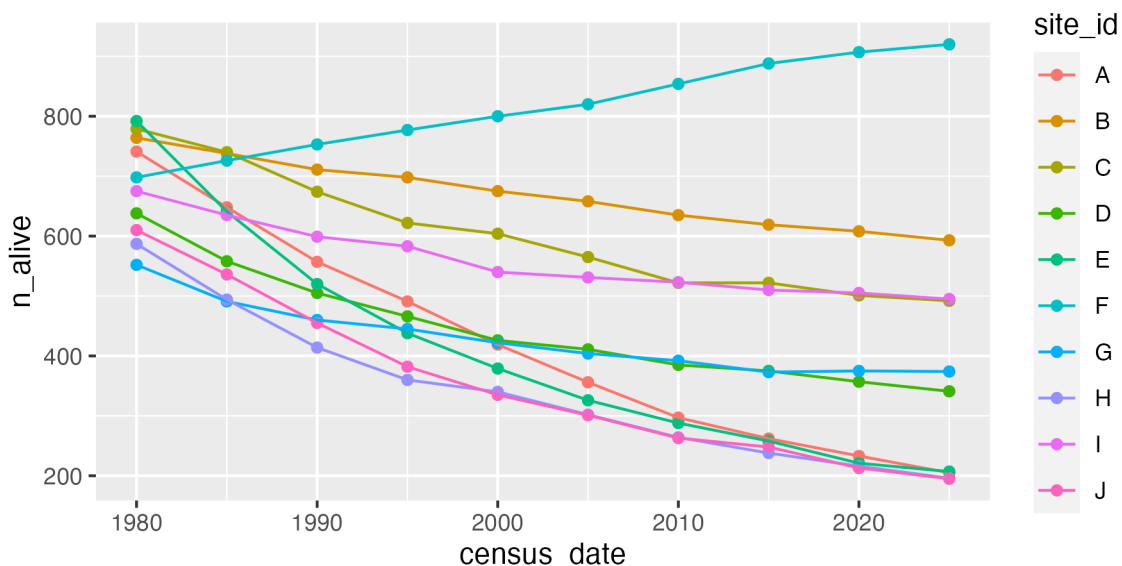


Figure 2: Number of stems over time for each site.

Adapt the code above to calculate the total biomass for each census per site and visualise it using `ggplot()`. Hint: use `sum(agb)` instead of `sum(diam != 0)`.

Which sites are decreasing in number of stems, which are increasing? Does this match the trends seen in biomass change over time?

We can also calculate the number of deaths recorded at each census. Recall that dead stems are recorded with a diameter of zero:

```
stems_mort_summ <- stems %>%
  group_by(site_id, census_date) %>%
  summarise(n_cum_mort = sum(diam == 0)) %>% # Cumulative mortality per census
  mutate(n_mort = n_cum_mort - lag(n_cum_mort)) %>% # Mortality per census
  dplyr::select(-n_cum_mort) # Remove intermediate column
```

Visualise the number of deaths over time by adapting the code you used above to visualise the number of living stems over time.

If we know the number of stems at each census, and the number of stems which died between each census, we can calculate the number of recruits recorded at each census, by re-arranging this simple equation:

$$N_t = N_{t-1} + R - M$$

Where N_t is the number of living stems at time t , N_{t-1} is the number of living stems at time $t - 1$ (i.e. the previous census), R is the number of recruits, and M is the number of deaths (mortality).

We can re-arrange the equation so it becomes:

$$R = N_t - N_{t-1} + M$$

Using `lag(n_alive)` to calculate N_{t-1} , calculate the number of recruits (R) at each census point, then plot the results using the same `ggplot()` code as above. Do recruitment rates appear to be increasing or decreasing over time? What simple parametric statistical test could you perform to check if there is a significant increase or decrease over time?

6 Calculating rates of production and loss

While the visualisations above give us some idea of the population and biomass dynamics of each site over the census period, to derive numbers which we can use to measure the carbon sink effect, and which we can use to parameterise an empirical model of biomass change, we need to calculate rates of production and loss.

Let's measure average biomass dynamics across the entire census period, 1980 to 2025, 45 years.

First, extract the first and last census:

```
stems_fl <- stems %>%
  filter(census_date %in% c(1980, 2025))
```

Then, measure the biomass at the first and last census for each site:

```
stems_fl_summ <- stems_fl %>%
  group_by(site_id, census_date) %>%
  summarise(agb = sum(agb)) %>%
  pivot_wider( # Transform the data from "long" format to "wide" format
    names_from = census_date,
    names_prefix = "agb_",
    values_from = agb)
```

We can now calculate biomass change across the census period for each site:

```
# Absolute biomass change (delta-AGB)
stems_fl_summ$dagb <- (stems_fl_summ$agb_1980 - stems_fl_summ$agb_2025)

# Mean biomass change per year
stems_fl_summ$dagby <- stems_fl_summ$dagb / (2025-1980)
```

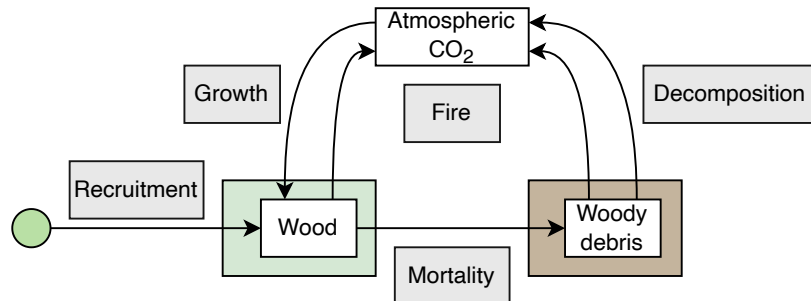


Figure 3: NPP-centric model of forest carbon fluxes.

Using the structural model of fluxes from the lecture (above), describe what part of the model the value of `dagb` represents.

Now, let's calculate rates of production and loss across the census period. Recall that we can calculate simple production and loss per year using the equations:

$$P_s = \frac{B_T - B_{s0}}{T}$$

$$L_s = \frac{B_0 - B_{s0}}{T}$$

where B_T is the biomass at the final census, B_{s0} is the biomass at the first census of the individuals which survived to the final census, B_0 is the biomass at the first census, and T is the length of the census period, in years.

Run the code below to calculate the various parameters for the equations for each site:

```
# Create a list, where each element is a dataframe containing a single site
stems_split <- split(stems, stems$site_id)

# Get the first census
first_census <- min(stems$census_date)

# Get the last census
final_census <- max(stems$census_date)

# For each site in the list:
params <- lapply(stems_split, function(x) {
  # Get the stem IDs for the stems which survived through the whole census period
  surv <- x[x$census_date == final_census & x$diam != 0, "stem_id"]

  # Find the biomass of the survivors at the first census
  Bs0 <- sum(x[x$census_date == first_census & x$stem_id %in% surv, "agb"])

  # Find the biomass at the final census
  BT <- sum(x[x$census_date == final_census, "agb"])

  # Find the biomass at the first census
```

```

B0 <- sum(x[x$census_date == first_census, "agb"])

# Get the census interval length
int <- final_census - first_census

# Create a dataframe for output
out <- data.frame(site_id = unique(x$site_id), Bs0, BT, B0, int)

# Return the dataframe
return(out)
})

# Join the dataframes together
params_df <- do.call(rbind, params)

```

Using the equations above and the parameters in `params_df`, calculate the simple production and loss for each site, and add as a column to `params_df`.

Referring back to the lecture slides, can you also calculate instantaneous rate of production (P) and loss (L). How do they differ from the “simple” estimates?

Using the rates of production and loss we can calculate the rate of net biomass change ($\delta B/\delta T$):

$$\delta B/\delta T = P_s - L_s$$

If $\delta B/\delta T > 0$ biomass is increasing. If $\delta B/\delta T < 0$ biomass is decreasing. Calculate $\delta B/\delta T$ for each site, does it match the visualisations of biomass over time?

Up to now, we have only calculated rates of biomass production and loss as an average across the entire census period. How might this affect our estimates of productivity?

Using a single site as an example, calculate production (P_s) and loss (L_s) using the periods 1980-2000, and 2000-2020, then calculate the mean of these rates across the two periods, how does it compare to the values calculated across the entire period?

Under what circumstances would a longer census interval (i.e. a longer time between censuses) lead to an under-estimate of total productivity?

7 Environmental drivers of biomass dynamics

Now that we have calculated some rates of productivity and loss (`params_df`), we can use simple statistical analyses to see if there is any relationship between biomass dynamics and environment.

First, we need to join the parameters dataframe to the sites dataframe, using the column `site_id` as a key:

```
sites_rates <- full_join(sites, params_df, by = "site_id")
```

Now, we can run linear models of the rates vs. the environmental drivers. For example:

```
mod_ex <- lm(Ps ~ bio19, data = sites_rates)
summary(mod_ex)
```

Use `ggplot()` to visualise the relationship between `bio19` and `Ps` and add a linear model line of best fit using `geom_smooth(method = "lm")`. Is there a strong relationship between these

variables? Why not? Hint: the sites are in the Amazon rainforest.

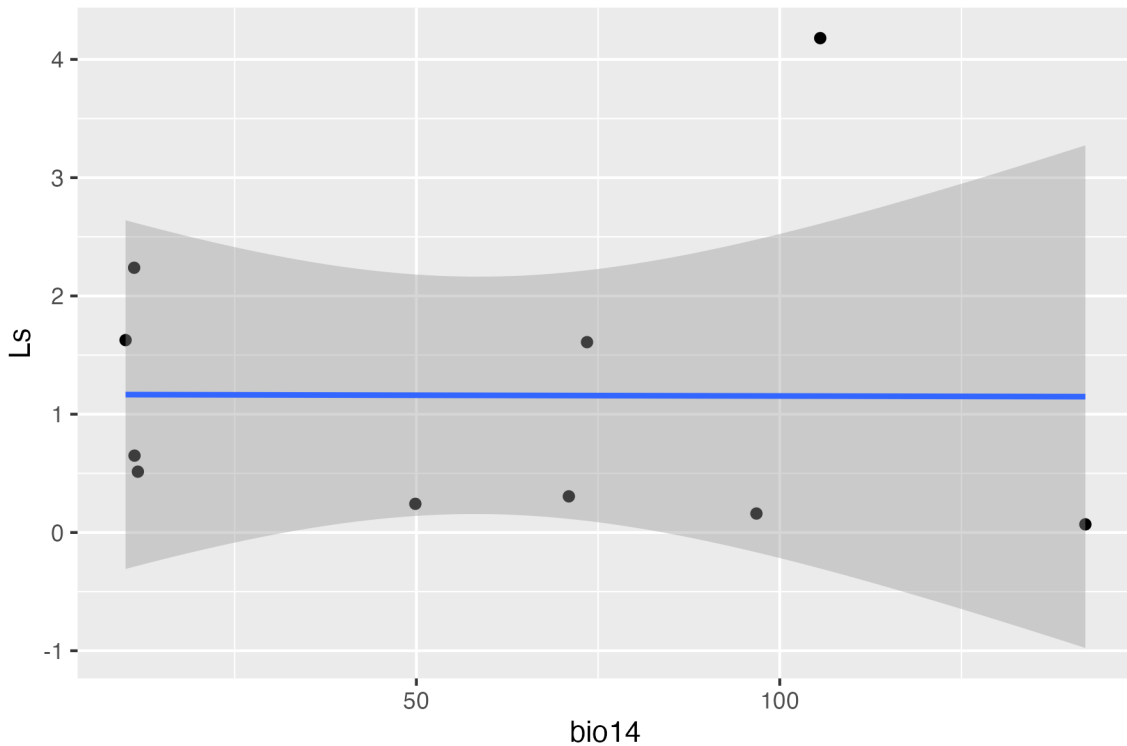


Figure 4: Example of plot showing relationship between bio14 and Ls.

Look through the list of BioClim variables on the first page of this worksheet. Pick some variables which you think might have an impact on some of the biomass metrics you have calculated above. Construct a hypothesis which predicts the direction of the effect of the variable on that metric. Run a linear model and inspect the model coefficients. Does the model support your hypothesis?

We can use the same simple linear models to predict the biomass of an unmeasured site, given it's environment.

First, you must calculate the mean period biomass (B_w) for all the sites in our dataset. This is a weighted measure of the average biomass observed in the plot over the course of the census period. It is calculated as:

$$B_w = \frac{B_T - B_0}{\log(\frac{B_T}{B_0})}$$

Imagine a site with a mean annual precipitation of 2000 mm per year. Using a linear model of **bio12** vs. mean period biomass (B_w) for each site in **stems**, predict the mean period biomass of the unmeasured site. Hint: **predict()** can be used to extract predicted values of a response variable from a model, by feeding it a dataframe with values of each explanatory variable in the model.

8 Predicting biomass dynamics at an unmeasured site

Load the files **unk_stems.csv**:


```
unk_stems <- read.csv("unk_stems.csv")
```

`unk_stems` contains a table, with the diameters, heights and species for stems recorded at a single site (Z) during a single census in 1990. The site is situated close to site D, only about 1 km away. This site was abandoned after the first census because it was difficult to access during the wet season.

Our goal is to predict what the biomass of this site might be in 2022.

First, calculate the biomass of each stem in 1990.

Then, using the intrinsic rate of change r for site D, extrapolate the biomass estimate of site Z forward to 2022.

Remember r is calculated as:

$$r = \ln\left(\frac{B_T}{B_0}\right)/T$$

We have essentially created an empirical model to describe the biomass dynamics of an unmeasured site, and parameterised that model using field data from a neighbouring site. What assumptions are we making in this model? List as many as you can.

All the rates of production and loss we have calculated up to now assume a linear process. If we use these rates to predict biomass change in an unmeasured system, the predicted biomass will continue to increase or decrease indefinitely, which is unrealistic over a long timescale. Using your understanding of forest successional processes and interactions among individuals in a population, think of some ways we might extend our model of forest growth to prevent this indefinite growth or decline.

9 Size-class and species specific rates of growth

In our previous model we assumed that the growth rates of all stems within a site were equal, using a single rate of biomass production calculated at the site level. This might be an over-simplification. We know that in tropical forests small stems in the understorey may grow very slowly, until they reach a certain size which allows them to compete with the larger canopy trees. Similarly, growth rates often decline when a tree is near death. We also know that tree species differ in their life-history strategies, with some trees growing quickly, often with the trade-off that they are at greater risk of death from drought or disease.

To check whether it is necessary to add size-class- and species-dependent growth rates to our model, we can measure the growth rates of individuals within different size classes and species.

Arbitrarily, we can calculate the growth rate for stems in these diameter size classes (cm):

- 5-10
- 10-20
- 20-40
- 40-60
- 60-80
- 80-100
- 100+

Remembering that we can calculate diameter growth over time (G_D) as:

$$G_D = \frac{D_t - D_{t-1}}{T}$$

We can first calculate the annual diameter growth rate for each stem across each census interval:

```
stems_d <- stems %>%
  filter(site_id == "D") %>% # Only keep site D
  arrange(census_date) %>% # Order the data by census date
  group_by(stem_id) %>%
  mutate(
    diam_g = diam - lag(diam), # Calculate diameter growth at each time point
```

```

census_int = census_date - lag(census_date), # Calculate census interval
diam_g_ann = diam_g / census_int) %>% # Annual measure of diam. growth
filter(
diam != 0, # Remove dead stems
!is.na(diam_g_ann)) # Remove first census, which has no growth measurement

```

Then “bin” our diameter data from site D by size class, using the `cut()` function:

```

stems_d$diam_cut <- cut(stems_d$diam, breaks = c(5,10,20,40,60,80,100,1000),
labels = c("5-10", "10-20", "20-40", "40-60", "60-80", "80-100", "100+"))

```

Now we can calculate the mean annual growth rate for each size class:

```

# Calculate mean diameter growth per size class
stems_d_g <- stems_d %>%
group_by(diam_cut) %>%
summarise(diam_g_ann_mean = mean(diam_g_ann))

```

We can then run a simple Analysis of Variance test (ANOVA) to check whether there are significant differences in annual diameter growth rates among size classes:

```

size_class_aov <- aov(diam_g_ann ~ diam_cut, data = stems_d)
summary(size_class_aov)

```

Is there a significant difference among size classes in their annual diameter growth rates?

Create a boxplot to visualise the distributions of diameter growth rates within each size class. Does the distribution match your understanding of how trees grow? Hint: use `geom_boxplot()` with `ggplot()` to create a boxplot.

Using the code above as a guide, investigate whether species differ in their annual diameter growth rates. Would it improve our model of biomass dynamics for site Z if we included species-specific growth rates? Hint: use `species_name` rather than `diam_cut`.