ECOSYSTEM SURVEILLANCE (AusPlots) TUTORIAL: UNDERSTANDING AND USING THE `ausplotsR` PACKAGE AND AusPlots DATA

This document contains a tutorial on how to access and use TERN's Ecosystem Surveillance (AusPlots) data.We will explore the use of both the package 'ausplotsR' & the 'ausplots' data that can be downloaded with this package.

`ausplotsR` is an R package for live extraction and preparation of TERN AusPlots ecosystem monitoring data. Through ausplotsR, users can: (1) directly obtain plot-based data on vegetation and soils across Australia, and (2) preprocess these data into structures that facilitate the visualisation and analysis of ausplots data. Data preprocessing includes the computation of species occurrence, vegetation cover, growth form, and basal area.

In this section of the workshop we will cover the following aspects:

1. ACCESSING AND INSTALLING THE `ausplotsR` PACKAGE (plus its Dependencies).

2. OBTAIN & EXPLORE AusPlots DATA: `get_ausplots` function:

- `get_ausplots` function
- Explore the structure of the obtained AusPlots data.

3. MANIPULATING AusPlots DATA:

- Find the 5 most sampled Bioregions.
- Subset sites in the 5 most sampled Bioregions (in all DFs in the list)

4. MAP THE SITES

- Obtain and prepare a map of Australia
- Plot AusPlots sites in the 5 most sampled Bioregions on the map of Australia.

5. SPECIES-LEVEL DATA: `species_table` function and species occurrence matrices (for the 5 most sampled Bioregions).

- First step: Create a species occurrence matrix. Compute Species by Site table using the function `species_table`
- Species Abundance/Percent Cover:
  - Percent Cover (Abundance) by Site Visit x Species (i.e. in all 'cells')
  - Abundance (Cover %) by Species. Find and plot 4 most abundant species on a map (dot size proportional to Abundance).
- Species Occurrence (Presence/Absence):
  - Presence/Absence across all Cells (i.e. Site Visit x Species).
  - Total Presence/Absence for each Species (i.e. per data frame Column):
    - Frequencies: Absolute and Relative.
    - Calculate and Plot (histogram): Presence (Absolute and Relative) Frequencies.
- Species Diversity:
  - Calculate various indices and create a data frame with these indices.
  - Plot 2 indices: Species Richness (from vouchers, more species recorded), and Shannon Index (from veg.PI for abundances) on a map (dot size proportional to the relevant diversity metric) .
- Rank-Abundance Curves (= Whittaker Plots) & Relative Abundance Models
  - Rank-Abundance Curves for the First 5 Site-Visits
  - Possible Models of Relative Abundance for one Community
  - Rank-Abundance Curves for each Bioregion (using the Species Mean Cover)

6. PROPORTIONAL VEGETATION COVER (= FRACTIONAL COVER): `fractional_cover' function.

- Latitudinal pattern in proportional vegetation cover (for a random subset of 200 sites).
- Temporal Variation in Fractional Cover: Explore, display, and assess (for 5 sites visited twice).

7. GROWTH FORM: `growth_form_table` function (for 5 most sampled bioregions)

- Plant Growth Forms Percent Cover against Sites : Compute using `growth_form_table`
- Cluster (Hierarchical Clustering) by Plant Growth Forms Percent Cover, colour branches by bioregion.

8. TOTAL VEGETATION COVER BY GROWTH FORM AND/OR HEIGHT: `single_cover_value` function (for 5 most sampled bioregions).

- Total Vegetation Cover of Any Green Vegetation >= 2m in Height
- Total Vegetation Cover of Trees >= 5m in Height (i.e. default arguments)
- Total Vegetation Cover of "Tussoc grass" of any Height (i.e. >= 0m height)

9. BASAL AREA (OR NUMBER OF BASAL WEDGE HITS): `basal_area` function (for 5 most sampled bioregions).

- Basal Area for each plot (m2/ha): Compute using `basal_area` .
- Display Basal Areas on map of Australia (dots size proportional to Basal Area).
- Boxplot of Basal Areas by Bioregion.

REQUIRED LIBRARIES

To run the R scripts in this tutorial a number of R packages (and their dependencies) must be installed. In addition to `ausplotsR` , which is at the core of this tutorial, the following packages are required: `dendextend` , `ggplot2` , `goeveg` , `gridExtra` , `mapdata` , `maps` , `maptools` , `sp` , and `vegan` .

The first step to install packages in R is selecting the CRAN (Comprehensive R Archive Network) mirror. Mirror selection and package installation can be done via R's menu (Packages/Set CRAN mirror… followed by Packages/install package(s)…) or programmatically the function `install.packages` (selecting the CRAN mirror using the argument `repos`). Typically is best to choose the `cloud` mirror (which automatically redirects to an appropriate server worldwide) or a mirror close to you (e.g. in your institution, country,..). A list of Comprehensive R Archive Network (CRAN) mirror URLs can be found here (https://cran.r-project.org/mirrors.html).

This is how you can install and load the R packages required for this tutorial.

If you need to install any of the required packages but `ausplotsR`, which is a special case (see below), uncomment the script below.

```
## Select the repository (i.e. CRAN mirror URL)
#my.repos = "https://cloud.r-project.org/"
#my.repos = "https://cran.csiro.au/"  # Example of an Australian mirror

## Install other required libraries
#install.packages(c("vegan", "goeveg", "maps", "maptools", "mapdata", "sp", ggplot2", "gridEx
tra", "ggspatial", "dendextend"), repos=my.repos)
```

Now the packages can be loaded using the `library` command.

```
# Load packages
library(ausplotsR)  # If not loaded above
library(vegan)
library(goeveg)

library(maps)
library(maptools)
library(mapdata)
library(sp)
library(ggplot2)
library(gridExtra)
#library(ggspatial)

library(dendextend)
```

ACCESSING AND INSTALLING THE `ausplotsR` PACKAGE (plus its Dependencies)

Currently `ausplotsR` must be installed directly from github using the 'devtools' package, which must have been previously installed. The GitHub site for `ausplotsR` contains the latest developments and information on the package; it can be found in this link (https://github.com/ternaustralia/ausplotsR).

```
## Install directly from github using the 'devtools' package
## Thus, 'devtools' must be previouly installed
install.packages("devtools", repos="https://cloud.r-project.org/")
```

```
## package 'devtools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\uqbblanc\AppData\Local\Temp\RtmpkpE8M3\downloaded_packages
```

```
library(devtools)
install_github("ternaustralia/ausplotsR", build_vignettes = TRUE)

## Load the package
library(ausplotsR)

## Obtaining Help and Initial Steps
help(ausplotsR)
browseVignettes(package="ausplotsR")
```

OBTAIN & EXPLORE AusPlots DATA: `get_ausplots` function

This function extracts and compiles AusPlots data.

Data of specific types, sites, geographical locations, and/or species can be requested via the function arguments.

DATA TYPES: Up to 8 different types of data can be obtained by setting the corresponding arguments to TRUE/FALSE:

- `site_info` : Site summary data. Includes (among others): plot and visit details, landform data, geographic coordinates, and notes. Included by default.
- `structural_summaries` : Site vegetation structural summaries
- `veg.vouchers` : Complete set of species records for the plot determined by a herbarium plus ID numbers for silica-dried tissue samples. Included by default.
- `veg.PI` : Point Intercept (PI) data. Includes data on: substrate, plant species, growth form and height, etc at each of (typically) 1010 points per plot. Included by default.
- `basal.wedge` : Basal Wedge Data Raw Hits. These data are required for the calculation of Basal Area by Species by Plot.
- `soil_subsites` : Information on what soil and soil metagenomics samples were taken at nine locations across the plot and their identification barcode numbers.
- `soil_bulk_density` :
- `soil_character` : Soil characterisation and sample ID data at 10 cm increments to a depth of 1 m.

SPATIAL FILTERING: AusPlot data can be spatially subset via the `get_ausplots` function arguments in two ways:

- `my.Plot_IDs` : Character vector with the plots IDs of specific AusPlots plots.
- `bounding_box` : Spatial filter for selecting AusPlots based on a rectangular box, in the format of e.g. c(xmin, xmax, ymin, ymax). AusPlots spatial data are are in longlat, thus x is the longitude and y is the latitude of the box/extent object (e.g., c(120, 140, -30, -10)).

SPECIES FILTERING: AusPlots data can also be subset by particular or sets of genus and/or species (i.e. as determined for the herbarium voucher) using the argument `species_name_search` . This optional argument takes the form of a character string indicating the terms to search and subset. Search terms are not case sensitive and do not require an exact taxonomic match (e.g. "Eucalyptus moderata", "Eucalyptus", and "euca" are all acceptable search terms). Species Filtering behaviour slightly differs among Data Types (i.e. for the different types of created Data Frames):

- For `veg.vouch` and `basal.wedge` , when these arguments are set to `TRUE` , `get_ausplots` returns data.frames with the corresponding data (i.e. voucher records and raw basal wedge data respectively) that match the species_name_search.
- For the remaing data types arguments, when these arguments are set to `TRUE` , `get_ausplots` returns data.frames with the corresponding data (e.g. point intercept data,…) for all plots where the species_name_search occurs.

The R object resulting from calling `get_ausplots` is a list of data frames containing the requested AusPlots data. The list includes a data frame for each type of data requested (i.e. up to 8 data frames: 'site_info', 'structural_summaries', …) and an auto-generated citation for the data extracted. Please cite ausplotsR and the TERN AusPlots data you use.

In each data frame the columns correspond to the variables supplied for each type of data and the number of rows (directly or indirectly) depends on the sites (i.e. via `my.Plot_IDs` or `bounding_box` if subsetted) or species (i.e. via `species_name_search` if subset) retrieved.

There are several variables common to all data frames. These include `site_location_name`, `site_location_visit_id`, and `site_unique` (a combination of the previous two). These variables can be used to interrelate the data frames. For example, the contents of two data frames can be combined using the common variable as a link (i.e. guidance to add the merged contents in the correct row). We will see multiple examples of data frame contents merges later in this tutorial. The variable 'site_unique' is typically the best option to link data frames, as it is the most specific variable representing a single visit to a particular site and it should be used in most analyses. Otherwise, errors such including data from the wrong visit to a site can occur.

```
# Example 1: All available data (i.e. all data types) for 3 plots
# ===============================================================

# Obtain the data ('site_info', 'veg.vouchers', and 'veg.PI' are retraived by default)
AP.data = get_ausplots( my.Plot_IDs=c("SATFLB0004", "QDAMGD0022", "NTASTU0002"),
                        structural_summaries=TRUE, basal.wedge=TRUE,
                                soil_subsites=TRUE, soil_bulk_density=TRUE, soil_character=
TRUE  )
```

```
## User-supplied Plot_IDs located.
```

```
# Explore retrieved data
class(AP.data)
```

```
## [1] "list"
```

```
summary(AP.data)
```

```
##                 Length Class      Mode
## site.info       43     data.frame list
## struct.summ     15     data.frame list
## soil.subsites   12     data.frame list
## soil.bulk       15     data.frame list
## soil.char       34     data.frame list
## veg.basal       10     data.frame list
## veg.vouch       12     data.frame list
## veg.PI          13     data.frame list
## citation        1      -none-     character
```

```
str(AP.data)
```

```
## List of 9
##  $ site.info    :'data.frame':   4 obs. of  43 variables:
##   ..$ site_location_name        : chr [1:4] "QDAMGD0022" "SATFLB0004" "SATFLB0004" "NTASTU
0002"
##   ..$ established_date           : chr [1:4] "2013-06-04T00:00:00" "2012-09-18T00:00:00" "2
012-09-18T00:00:00" "2016-05-01T16:58:00"
##   ..$ description                : chr [1:4] "Mackunda Downs Station, 500m east of homestea
d.  26km west of Middleton." "Brachina Gorge Heysen Range Lower. 63km North North East of Ade
laide" "Brachina Gorge Heysen Range Lower. 63km North North East of Adelaide" "Maryfield Stat
ion, 7.6km north north west of homestead. 27.5km south east of Larrimah"
##   ..$ bioregion_name            : chr [1:4] "MGD" "FLB" "FLB" "STU"
##   ..$ landform_pattern          : chr [1:4] "ALP" "MOU" "MOU" "PLA"
##   ..$ landform_element          : chr [1:4] "PLA" "HSL" "HSL" "PLA"
##   ..$ site_slope                : chr [1:4] "1" "17" "17" "0"
##   ..$ site_aspect               : chr [1:4] "180" "225" "225" NA
##   ..$ comments                  : chr [1:4] "Astrebla pectinata / Cenchrus ciliaris / Astr
ebla elymoides low open tussock grassland on alluvial plain adjoin"| __truncated__ "Grazing i
mpact high- goat tracks and droppings. Rabbit droppings also. Lots of Yellow footed rock wall
abies clo"| __truncated__ "Grazing impact high- goat tracks and droppings. Rabbit droppings a
lso. Lots of Yellow footed rock wallabies clo"| __truncated__ "Plot is flat. Low mound ( Like
ly anthropogenic) made up of ironstone gravels at the north west corner. Minimal "| __truncat
ed__
##   ..$ outcrop_lithology         : chr [1:4] "NA" "SA" "SA" "NA"
##   ..$ other_outcrop_lithology   : chr [1:4] "NA" "NC" "NC" "NC"
##   ..$ plot_dimensions           : chr [1:4] "100m x 100m." NA NA NA
##   ..$ site_location_visit_id    : int [1:4] 53501 58658 53705 58429
##   ..$ visit_start_date          : chr [1:4] "2013-05-18T09:34:00" "2018-08-28T09:20:29" "2
012-09-18T00:00:00" "2016-05-01T16:58:00"
##   ..$ visit_end_date            : chr [1:4] "2013-05-18T09:34:00" "2018-08-28T09:20:29" "2
012-09-18T00:00:00" "2016-05-01T16:58:00"
##   ..$ visit_notes               : chr [1:4] "" NA NA "Corymbia polycarpa and Corymbia term
inalis combined for Basal area\r\n\r\nunknown substrate in point intercept "| __truncated__
##   ..$ location_description      : chr [1:4] "Mackunda Station, north of Middleton." "Brach
ina lower.Largely unchanged since previous visit possibly more Carrichtera annua. Grazing imp
act goat, rab"| __truncated__ "Brachina Gorge Heysen Range Lower. 63km North North East of Ad
elaide.Grazing impact high- goat tracks and dropp"| __truncated__ "Maryfield Station, 7.6km n
orth north west of homestead. 27.5km south east of Larimah"
##   ..$ erosion_type              : chr [1:4] "G" "n/a" "n/a" "n/a"
##   ..$ erosion_abundance         : chr [1:4] "2" "X" "1" "X"
##   ..$ erosion_state             : chr [1:4] "NC" "n/a" "n/a" "n/a"
##   ..$ microrelief               : chr [1:4] "Z" "NH" "NC" "TM"
##   ..$ drainage_type             : int [1:4] 4 3 6 4
##   ..$ disturbance               : chr [1:4] "1L" "1L" "1H" "0"
##   ..$ climatic_condition        : chr [1:4] "DRY" "DRY" "WET" "DRY"
##   ..$ vegetation_condition      : chr [1:4] "DRY" "AVG" "AVG" "AVG"
##   ..$ observer_veg              : int [1:4] 3 1 16 1
##   ..$ observer_soil             : int [1:4] 2 31 1 2
##   ..$ described_by              : int [1:4] 3 1 16 1
##   ..$ pit_marker_easting        : num [1:4] 529568 268527 268528 326265
##   ..$ pit_marker_northing       : num [1:4] 7526350 6531518 6531576 8256078
##   ..$ pit_marker_mga_zones      : int [1:4] 54 54 53 53
##   ..$ pit_marker_datum          : chr [1:4] "WGS84" "WGS84" "GDA94" "WGS84"
##   ..$ pit_marker_location_method: chr [1:4] "GPS" NA "GPS" NA
##   ..$ soil_observation_type     : chr [1:4] "P" "PA" "P" "P"
##   ..$ a_s_c                     : chr [1:4] "NC" NA NA NA
##   ..$ plot_is_100m_by_100m      : logi [1:4] TRUE TRUE TRUE TRUE
##   ..$ plot_is_aligned_to_grid   : logi [1:4] TRUE TRUE TRUE TRUE
```

```
##    ..$ plot_is_permanently_marked: logi [1:4] TRUE TRUE TRUE TRUE
##    ..$ latitude                  : num [1:4] -22.4 -31.3 -31.3 -15.8
##    ..$ longitude                 : num [1:4] 141 139 139 133
##    ..$ point                     : chr [1:4] "SW" "SW" "SW" "SW"
##    ..$ state                     : chr [1:4] "QLD" "SA" "SA" "NT"
##    ..$ site_unique               : chr [1:4] "QDAMGD0022-53501" "SATFLB0004-58658" "SATFLB0
004-53705" "NTASTU0002-58429"
##  $ struct.summ  :'data.frame':   4 obs. of  15 variables:
##    ..$ site_location_name    : chr [1:4] "QDAMGD0022" "SATFLB0004" "NTASTU0002" "SATFLB000
4"
##    ..$ site_location_visit_id: int [1:4] 53501 53705 58429 58658
##    ..$ phenology_comment     : chr [1:4] "" "Ptilotus obovatus var. obovatus  flowering. Tr
iodia sp. has no seeds. No fruit on Callitris glaucophylla - no e"| __truncated__ "NC" "NC"
##    ..$ upper_1_dominant      : chr [1:4] "" "SAT 000251" "NTA017194" "SAT007365"
##    ..$ upper_2_dominant      : chr [1:4] "" "SAT 000229" "NTA017232" "SAT007407"
##    ..$ upper_3_dominant      : chr [1:4] "" NA "NTA017084" "None"
##    ..$ mid_1_dominant        : chr [1:4] "" "SAT 000244" NA "SAT007387"
##    ..$ mid_2_dominant        : chr [1:4] "" "SAT 000261" NA "SAT007423"
##    ..$ mid_3_dominant        : chr [1:4] "" NA NA "SAT007411"
##    ..$ ground_1_dominant     : chr [1:4] "QDA  003325" "SAT 000233" "NTA017070" "SAT007375"
##    ..$ ground_2_dominant     : chr [1:4] "QDA  003293" NA "NTA017076" "SAT007353"
##    ..$ ground_3_dominant     : chr [1:4] "QDA  003325" NA "NTA017082" "SAT007419"
##    ..$ description           : chr [1:4] "Astrebla pectinata / Cenchrus ciliaris / Astrebla
elymoides low open tussock grassland with scattered ." "Callitris glaucophylla / Eucalyptus i
ntertexta low woodland. A mid-stratum dominated by Rhagodia paradoxa and H"| __truncated__ "C
orymbia terminalis mixed mid woodland with  Corymbia polycarpa / Eucalyptus pruinosa/ Eucalyp
tus chlorophylla "| __truncated__ "Eucalyptus intertexta low open woodland  with Callitris gl
aucophylla.  A mid-storey of Rhagodia parabolica,  Do"| __truncated__
##    ..$ mass_flowering_event  : logi [1:4] FALSE FALSE FALSE FALSE
##    ..$ site_unique           : chr [1:4] "QDAMGD0022-53501" "SATFLB0004-53705" "NTASTU0002-
58429" "SATFLB0004-58658"
##  $ soil.subsites:'data.frame':   36 obs. of  12 variables:
##    ..$ site_location_name     : chr [1:36] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "QDAMGD0
022" ...
##    ..$ site_location_visit_id : int [1:36] 53501 53501 53501 53501 53501 53501 53501 53501
53501 53705 ...
##    ..$ subsite_id             : chr [1:36] "1" "2" "3" "4" ...
##    ..$ zone                   : int [1:36] 54 54 54 54 54 54 54 54 54 53 ...
##    ..$ easting                : int [1:36] 529581 529580 529582 529592 529595 529620 52966
4 529587 529663 268530 ...
##    ..$ northing               : int [1:36] 7526343 7526338 7526335 7526338 7526345 7526378
7526417 7526423 7526333 6531529 ...
##    ..$ ten_to_twenty_barcode  : chr [1:36] "QDA  051589" "QDA  051592" "QDA  051595" "Q
DA  051598" ...
##    ..$ zero_to_ten_barcode    : chr [1:36] "QDA  051588" "QDA  051591" "QDA  051594" "Q
DA  051597" ...
##    ..$ twenty_to_thirty_barcode: chr [1:36] "QDA  051590" "QDA  051593" "QDA  051596" "Q
DA  051599" ...
##    ..$ comments               : chr [1:36] "bare ground" "between grass tussocks" "between
grass tussocks" "bare ground" ...
##    ..$ metagenomic_barcode    : chr [1:36] "QDA 053721" "QDA 053722" "QDA 053723" "QDA 053
724" ...
##    ..$ site_unique            : chr [1:36] "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD00
22-53501" "QDAMGD0022-53501" ...
##  $ soil.bulk    :'data.frame':   9 obs. of  15 variables:
##    ..$ site_location_name    : chr [1:9] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "NTASTU00
02" ...
##    ..$ site_location_visit_id : int [1:9] 53501 53501 53501 58429 58429 58429 58658 58658
```

```
58658
##    ..$ sample_id            : chr [1:9] "0" "1" "2" "0" ...
##    ..$ paper_bag_weight     : logi [1:9] NA NA NA NA NA NA ...
##    ..$ oven_dried_weight_in_bag: logi [1:9] NA NA NA NA NA NA ...
##    ..$ ring_weight          : logi [1:9] NA NA NA NA NA NA ...
##    ..$ gravel_weight        : int [1:9] NA NA NA 0 0 0 NA NA NA
##    ..$ ring_volume          : num [1:9] NA NA NA 209 209 ...
##    ..$ gravel_volume        : int [1:9] NA NA NA 0 0 0 NA NA NA
##    ..$ fine_earth_weight_in_bag: int [1:9] NA NA NA 0 0 0 0 0
##    ..$ fine_earth_weight    : int [1:9] NA NA NA 0 0 0 0 0
##    ..$ fine_earth_volume    : num [1:9] NA NA NA 209 209 ...
##    ..$ fine_earth_bulk_density : int [1:9] NA NA NA 0 0 0 0 0
##    ..$ gravel_bulk_density  : int [1:9] NA NA NA 0 0 0 0 0
##    ..$ site_unique          : chr [1:9] "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD002
2-53501" "NTASTU0002-58429" ...
## $ soil.char   :'data.frame':   28 obs. of  34 variables:
##    ..$ site_location_name   : chr [1:28] "SATFLB0004" "QDAMGD0022" "QDAMGD0022" "QDAMGD002
2" ...
##    ..$ site_location_visit_id: int [1:28] 53705 53501 53501 53501 53501 53705 53501 53705 5
3501 53501 ...
##    ..$ upper_depth          : num [1:28] 0 0.5 0.8 0.6 0.7 0.1 0.3 0.6 0.2 0 ...
##    ..$ lower_depth          : num [1:28] 0.1 0.6 0.9 0.7 0.8 0.2 0.4 0.7 0.3 0.1 ...
##    ..$ horizon              : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ texture_grade        : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ texture_qualifier    : chr [1:28] NA "NC" "NC" "NC" ...
##    ..$ texture_modifier     : chr [1:28] NA "NC" "NC" "NC" ...
##    ..$ colour_when_moist    : chr [1:28] NA "NC" "NC" "NC" ...
##    ..$ colour_when_dry      : chr [1:28] NA "NC" "NC" "NC" ...
##    ..$ mottles_colour       : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ mottles_abundance    : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ mottles_size         : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ segregations_abundance: chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ segregations_size    : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ segregations_nature  : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ segregations_form    : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ comments             : chr [1:28] NA NA NA NA ...
##    ..$ collected_by         : int [1:28] 8 4 4 4 4 NA 4 NA 4 4 ...
##    ..$ smallest_size_1      : chr [1:28] "11" "11" "11" "11" ...
##    ..$ smallest_size_2      : logi [1:28] NA NA NA NA NA NA ...
##    ..$ effervescence        : chr [1:28] "N" "N" "N" "N" ...
##    ..$ ec                   : num [1:28] 0.07 0.18 0.54 0.38 0.52 0.03 0.07 1.84 0.06 0.04
...
##    ..$ ph                   : num [1:28] 6.3 8.3 8.1 8.3 8.1 7 8 8.3 7.9 7.8 ...
##    ..$ pedality_grade       : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ pedality_fabric      : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ next_size_type_2     : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ next_size_type_1     : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ smallest_size_type_2 : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ smallest_size_type_1 : chr [1:28] "NC" "NC" "NC" "NC" ...
##    ..$ next_size_2          : logi [1:28] NA NA NA NA NA NA ...
##    ..$ next_size_1          : chr [1:28] "11" "11" "11" "11" ...
##    ..$ layer_barcode        : chr [1:28] "SAT005230" "QDA   051583" "QDA   051586" "QDA
051584" ...
##    ..$ site_unique          : chr [1:28] "SATFLB0004-53705" "QDAMGD0022-53501" "QDAMGD0022
-53501" "QDAMGD0022-53501" ...
## $ veg.basal   :'data.frame':   95 obs. of  10 variables:
##    ..$ site_location_name   : chr [1:95] "NTASTU0002" "SATFLB0004" "SATFLB0004" "SATFLB00
04" ...
```

```
##    ..$ site_location_visit_id : int [1:95] 58429 53705 53705 53705 53705 53705 53705 53705
53705 53705 ...
##    ..$ site_location_id      : int [1:95] 61138 60122 60122 60122 60122 60122 60122 60122
60122 60122 ...
##    ..$ point_id              : chr [1:95] "NE" "W" "SW" "SE" ...
##    ..$ herbarium_determination: chr [1:95] "Dead Tree/Shrub" "Alectryon oleifolius" "Alectr
yon oleifolius" "Alectryon oleifolius" ...
##    ..$ veg_barcode           : chr [1:95] "NO_BARCODE_DEAD_TREE_804159" "SAT 000242" "SAT
000242" "SAT 000242" ...
##    ..$ hits                  : int [1:95] 1 1 3 4 1 1 1 20 17 6 ...
##    ..$ basal_area_factor     : num [1:95] 0.1 0.25 0.5 0.5 0.5 0.25 0.5 0.25 0.5 0.5 ...
##    ..$ basal_area            : num [1:95] 0.1 0.25 1.5 2 0.5 0.25 0.5 5 8.5 3 ...
##    ..$ site_unique           : chr [1:95] "NTASTU0002-58429" "SATFLB0004-53705" "SATFLB000
4-53705" "SATFLB0004-53705" ...
##  $ veg.vouch   :'data.frame':   208 obs. of  12 variables:
##    ..$ site_location_name       : chr [1:208] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "QDAM
GD0022" ...
##    ..$ veg_barcode              : chr [1:208] "QDA  003331" "NO_BARCODE_FORB_950413164" "N
O_BARCODE_GRASS_656236361" "NO_BARCODE_DEAD_TREE_558409020" ...
##    ..$ herbarium_determination  : chr [1:208] "Glinus lotoides" "Annual forb" "Annual gras
s" "Dead tree/shrub" ...
##    ..$ is_uncertain_determination: logi [1:208] FALSE NA NA NA NA NA ...
##    ..$ visit_start_date         : chr [1:208] "2013-05-18T09:34:00" "2013-05-18T09:34:00"
"2013-05-18T09:34:00" "2013-05-18T09:34:00" ...
##    ..$ site_location_visit_id   : int [1:208] 53501 53501 53501 53501 53705 53705 53705 53
705 58429 58429 ...
##    ..$ primary_gen_barcode      : chr [1:208] "QDA  003332" NA NA NA ...
##    ..$ secondary_gen_barcode_1  : chr [1:208] NA NA NA NA ...
##    ..$ secondary_gen_barcode_2  : chr [1:208] NA NA NA NA ...
##    ..$ secondary_gen_barcode_3  : chr [1:208] NA NA NA NA ...
##    ..$ secondary_gen_barcode_4  : chr [1:208] NA NA NA NA ...
##    ..$ site_unique              : chr [1:208] "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMG
D0022-53501" "QDAMGD0022-53501" ...
##  $ veg.PI      :'data.frame':   4323 obs. of  13 variables:
##    ..$ site_location_name    : chr [1:4323] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "QDAMGD
0022" ...
##    ..$ site_location_visit_id : int [1:4323] 53501 53501 53501 53501 53501 53501 53501 5350
1 53501 53501 ...
##    ..$ transect              : Factor w/ 15 levels "E2-W2","E4-W4",..: 13 13 13 13 13 13 1
3 13 13 13 ...
##    ..$ point_number          : int [1:4323] 27 0 1 3 4 5 6 7 8 9 ...
##    ..$ veg_barcode           : chr [1:4323] NA NA NA "QDA  003325" ...
##    ..$ herbarium_determination: chr [1:4323] NA NA NA "Astrebla pectinata" ...
##    ..$ substrate             : chr [1:4323] "Bare" "Litter" "Bare" "Litter" ...
##    ..$ in_canopy_sky         : logi [1:4323] NA NA NA FALSE NA FALSE ...
##    ..$ dead                  : logi [1:4323] NA NA NA FALSE NA FALSE ...
##    ..$ growth_form           : chr [1:4323] NA NA NA "Tussock grass" ...
##    ..$ height                : num [1:4323] NA NA NA 0.2 NA 0.2 NA NA NA 0.1 ...
##    ..$ hits_unique           : chr [1:4323] "W1-E1 27" "W1-E1 0" "W1-E1 1" "W1-E1 3" ...
##    ..$ site_unique           : chr [1:4323] "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD0
022-53501" "QDAMGD0022-53501" ...
##  $ citation    : chr "TERN (2019) AusPlots ecosystem surveillance monitoring dataset (UR
L: http://aekos.org.au/collection/adelaide.ed"| __truncated__
```

```
# Example 2: Default data for a particular Geographic Extent
# ==========================================================

# 'site_info', 'veg.vouchers', and 'veg.PI' data retrived for Brisbane (27.4698S, 153.0251E)
# and its sourrounding area
AP.data = get_ausplots(bounding_box=c(152.5, 153.5, -28, -27))

# Explore retrieved data
#class(AP.data)    # As in Example 1 (can run uncommented if curious)
summary(AP.data)
```

```
##              Length Class      Mode
## site.info 43      data.frame list
## veg.vouch 12      data.frame list
## veg.PI    13      data.frame list
## citation   1      -none-     character
```

```
#str(AP.data)    # Similar to Example 1 (can run uncommented if curious)
```

```
# Example 3: 'Default data' + 'basal.wedge' + 'structural_summaries' for the genus Eucalyptus
# ===========================================================================================

# Default data frames ('site_info', 'veg.vouchers', and 'veg.PI') + 'basal.wedge' + structura
l_summaries data frames for the genus Eucalyptus
AP.data = get_ausplots(basal.wedge=TRUE, structural_summaries=TRUE, species_name_search="Euca
lyptus")

# Explore retrieved data
#class(AP.data)    # As in Example 1 (can run uncommented if curious)
summary(AP.data)
```

```
##                Length Class      Mode
## site.info   43      data.frame list
## struct.summ 15      data.frame list
## veg.basal   10      data.frame list
## veg.vouch   12      data.frame list
## veg.PI      13      data.frame list
## citation     1      -none-     character
```

```
#str(AP.data)    # Similar to Example 1 (can run uncommented if curious)

# Explore species contained in each data frame
head(AP.data$veg.vouch) # Includes Records that match 'eucalyptus'
```

```
##   site_location_name veg_barcode          herbarium_determination
## 1         QDAMUL0003  QDA 001432 Eucalyptus crebra x e. melanophloia
## 2         SASMDD0002  SAS 000461                  Eucalyptus oleosa
## 3         SASMDD0002  SAS 000462                  Eucalyptus dumosa
## 4         SASMDD0002  SAS 000463 Eucalyptus socialis subsp. socialis
## 5         SASMDD0002  SAS 000038     Eucalyptus oleosa subsp. oleosa
## 6         SASMDD0002  SAS 000039                  Eucalyptus dumosa
##   is_uncertain_determination   visit_start_date site_location_visit_id
## 1                      FALSE 2013-04-26T00:00:00                  53595
## 2                      FALSE 2012-09-23T00:00:00                  53711
## 3                      FALSE 2012-09-23T00:00:00                  53711
## 4                      FALSE 2012-09-23T00:00:00                  53711
## 5                      FALSE 2012-09-23T00:00:00                  53711
## 6                      FALSE 2012-09-23T00:00:00                  53711
##   primary_gen_barcode secondary_gen_barcode_1 secondary_gen_barcode_2
## 1               <NA>                    <NA>                    <NA>
## 2       not collected                    <NA>                    <NA>
## 3       not collected                    <NA>                    <NA>
## 4       not collected                    <NA>                    <NA>
## 5         SAS  000521            SAS  000522            SAS  000523
## 6         SAS  000526            SAS  000528            SAS  000529
##   secondary_gen_barcode_3 secondary_gen_barcode_4       site_unique
## 1                    <NA>                    <NA> QDAMUL0003-53595
## 2                    <NA>                    <NA> SASMDD0002-53711
## 3                    <NA>                    <NA> SASMDD0002-53711
## 4                    <NA>                    <NA> SASMDD0002-53711
## 5            SAS  000524            SAS  000525 SASMDD0002-53711
## 6            SAS  000527                    <NA> SASMDD0002-53711
```

```
head(AP.data$veg.PI)  # Includes Plots where 'eucalyptus' occurs
```

```
##   site_location_name site_location_visit_id transect point_number
## 1         WAACOO0006                  53438    S1-N1            0
## 2         WAACOO0006                  53438    S1-N1            1
## 3         WAACOO0006                  53438    S1-N1            2
## 4         WAACOO0006                  53438    S1-N1            3
## 5         WAACOO0006                  53438    S1-N1            4
## 6         WAACOO0006                  53438    S1-N1            5
##   veg_barcode            herbarium_determination substrate
## 1       <NA>                              <NA>    Crypto
## 2       <NA>                              <NA>    Crypto
## 3 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
## 4 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
## 5 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
## 6 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
##   in_canopy_sky  dead growth_form height hits_unique       site_unique
## 1            NA    NA       <NA>     NA    S1-N1 0 WAACOO0006-53438
## 2            NA    NA       <NA>     NA    S1-N1 1 WAACOO0006-53438
## 3         FALSE FALSE      Shrub    3.5    S1-N1 2 WAACOO0006-53438
## 4         FALSE FALSE      Shrub    3.5    S1-N1 3 WAACOO0006-53438
## 5         FALSE FALSE      Shrub    3.7    S1-N1 4 WAACOO0006-53438
## 6         FALSE FALSE      Shrub    3.5    S1-N1 5 WAACOO0006-53438
```

```
head(AP.data$veg.basal) # Includes Records that match 'eucalyptus'
```

```
##   site_location_name site_location_visit_id site_location_id point_id
## 1         WAAC000006                  53438            59857       NW
## 2         WAAC000006                  53438            59857       NW
## 3         WAAC000006                  53438            59857       NW
## 4         WAAC000006                  53438            59857        N
## 5         WAAC000006                  53438            59857        N
## 6         WAAC000006                  53438            59857        N
##   herbarium_determination veg_barcode hits basal_area_factor basal_area
## 1     Eucalyptus moderata WAA  001048    1               0.1       0.10
## 2     Eucalyptus salubris WAA  001083    5               0.1       0.50
## 3     Eucalyptus salubris WAA  001093    7               0.1       0.70
## 4     Eucalyptus moderata WAA  001048    0               0.1       0.00
## 5     Eucalyptus salubris WAA  001083    2               0.1       0.15
## 6     Eucalyptus salubris WAA  001093    4               0.1       0.40
##          site_unique
## 1 WAAC000006-53438
## 2 WAAC000006-53438
## 3 WAAC000006-53438
## 4 WAAC000006-53438
## 5 WAAC000006-53438
## 6 WAAC000006-53438
```

```
head(AP.data$struct.summ)  # Includes Plots where 'eucalyptus' occurs
```

```
##   site_location_name site_location_visit_id
## 1         QDAMUL0003                  53595
## 2         SASMDD0002                  53711
## 3         SASMDD0016                  57000
## 4         NSAMDD0005                  56969
## 5         QDAMUL0001                  53594
## 6         NTAGFU0032                  53679
##
phenology_comment
## 1  Mulga have just finished flowering but no fruit. Tussock grasses mostly dry. Dom hibisc
us in ground layer has just finished fruiting throughout the site
## 2
None
## 3
NC
## 4
NC
## 5
<NA>
## 6                                          Melaleuca stenostachya has finished fl
owering and with not much fruit present. Tussock grasses all dry.
##   upper_1_dominant upper_2_dominant upper_3_dominant mid_1_dominant
## 1       QDA 001428       QDA 001432                      QDA 001428
## 2       SAS 000461       SAS 000463       SAS 000462     SAS 000041
## 3        SAS001764        SAS001732             None      SAS001758
## 4       NSA 010375       NSA 010439       NSA 010391     NSA 010387
## 5       QDA 001355       QDA 001363       QDA 001329     QDA 001355
## 6                                                        NTA 004067
##   mid_2_dominant mid_3_dominant ground_1_dominant ground_2_dominant
## 1                                      QDA 001438        QDA 001402
## 2     SAS 000047     SAS 000049        SAS 000453        SAS 000465
## 3           None           None         SAS001770                -1
## 4           None           None        NSA 010433        NSA 010453
## 5           <NA>           <NA>        QDA 001341              <NA>
## 6     NTA 004077                       NTA 004037        NTA 004021
##   ground_3_dominant
## 1
## 2              <NA>
## 3                -1
## 4       NSA 010465
## 5              <NA>
## 6
##
description
## 1                                       Acacia aneura var. major low open forest w
ith emergent Eucalyptus crebra x E. melanophloia. Mid stratum of juvenile Acacia aneura var.
major. Sparse ground stratum dominated by Thyridolepis xerophila with Hibiscus sturtii
## 2                                       Eucalyptus oleosa / Eucalyptus socialis su
bsp. socialis / Eucalyptus dumosa mixed mid open Mallee forest. Sparse mid layer of isolated
shrubs and a sparse ground layer dominated by Maireana pentatropis and Austrostipa sp.
## 3 Eucalyptus oleosa  mid Mallee woodland with Eucalyptus gracilis. Mid stratum of Senna ar
temisioides subsp. coriacea and a ground stratum dominated by Zygophyllum aurantiacum subsp.
aurantiacum and mixed scattered chenopods mainly  Atriplex stipitata and Maireana spp.
## 4              Eucalyptus dumosa subsp. dumosa, E. socialis and E. gracilis Mallee low
woodland (4-6m) with mixed species mid layer dominated by Eremophila longifolia (1-3m) and gr
ound layer of Enchylaena tomentosa, Zygophyllum sp. and Maireana pentatropis on dunes.
## 5                                                     Acacia aneura var. major
```

wood low open forest with emergent Eucalyptus populnea and Eucalyptus melanophloia - E. white
i intergrade. Mid stratum of juvenile Acacia aneura var. major. Minimal ground stratum
6   Melaleuca stenostachya tall open shrubland with
a tussock grass ground stratum of Sorghum plumosum. and Aristida holathera. Some recruitment
in the Eucalyptus pruinosa and Melaleuca stenostachya with the cohort averaging 30cm.
mass_flowering_event   site_unique
1   FALSE QDAMUL0003-53595
2   FALSE SASMDD0002-53711
3   FALSE SASMDD0016-57000
4   FALSE NSAMDD0005-56969
5   FALSE QDAMUL0001-53594
6   FALSE NTAGFU0032-53679

```
# Example 4: 'site_info', 'veg.PI', and 'basal.wedge' data for all sites
# =====================================================================

# Retreive data
start.time = Sys.time()
AP.data = get_ausplots(veg.vouchers=FALSE, basal.wedge=TRUE)
end.time = Sys.time()
end.time - start.time
```

Time difference of 41.76656 secs

```
# Explore
#class(AP.data)   # As in Example 1 (can run uncommented if curious)
summary(AP.data)
```

```
##            Length Class      Mode
## site.info 43     data.frame list
## veg.basal 10     data.frame list
## veg.PI    13     data.frame list
## citation   1     -none-     character
```

```
#str(AP.data)   # Similar to Example 1 (can run uncommented if curious)

# Explore 'site_info' data
dim(AP.data$site.info)
```

[1] 662 43

```
names(AP.data$site.info)
```

```
##  [1] "site_location_name"        "established_date"
##  [3] "description"               "bioregion_name"
##  [5] "landform_pattern"          "landform_element"
##  [7] "site_slope"                "site_aspect"
##  [9] "comments"                  "outcrop_lithology"
## [11] "other_outcrop_lithology"   "plot_dimensions"
## [13] "site_location_visit_id"    "visit_start_date"
## [15] "visit_end_date"            "visit_notes"
## [17] "location_description"      "erosion_type"
## [19] "erosion_abundance"         "erosion_state"
## [21] "microrelief"               "drainage_type"
## [23] "disturbance"               "climatic_condition"
## [25] "vegetation_condition"      "observer_veg"
## [27] "observer_soil"             "described_by"
## [29] "pit_marker_easting"        "pit_marker_northing"
## [31] "pit_marker_mga_zones"      "pit_marker_datum"
## [33] "pit_marker_location_method" "soil_observation_type"
## [35] "a_s_c"                     "plot_is_100m_by_100m"
## [37] "plot_is_aligned_to_grid"   "plot_is_permanently_marked"
## [39] "latitude"                  "longitude"
## [41] "point"                     "state"
## [43] "site_unique"
```

```
head(AP.data$site.info)
```

```
##   site_location_name   established_date
## 1         WAANUL0007 2014-09-06T15:24:41
## 2         NTAFIN0031 2012-10-25T00:00:00
## 3         QDAMUL0003 2013-04-26T00:00:00
## 4         NTAFIN0004 2011-10-06T00:00:00
## 5         NTAFIN0004 2011-10-06T00:00:00
## 6         SASMDD0002 2012-09-23T00:00:00
##                                                               description
## 1         Great Victoria Desert Nature Reserve, 102.2km south east of Tjuntjuntjara
## 2 Umbeara Station 26.5km South East of Umbeara Homestead. 11km North of SA/Not Border
## 3                                            61km SE of Issiford on Idalia NP
## 4        Top of James Range, Owen Springs Reserve . 73km South West of Alice Springs
## 5        Top of James Range, Owen Springs Reserve . 73km South West of Alice Springs
## 6                              Calperum Station, 23km North West of Renmark
##   bioregion_name landform_pattern landform_element site_slope site_aspect
## 1            NUL              PLA              DDE       <NA>        <NA>
## 2            FIN              LOW              HSL       <NA>        <NA>
## 3            MUL              PLT              HSL          4         225
## 4            MAC              HIL              HCR          8         135
## 5            MAC              HIL              HCR          8         135
## 6            MDD              LON              DUN          0        <NA>
##
comments
## 1 Plot is flat but sits in a drainage depression between very low rises. Some limestone co
bbles and gravel- larger ones with cryptogam crust. Some very low limestone outcrop- almost a
t ground level. Very long unburnt but difficult to tell exactly how long. Grazing effect is l
ow- some evidence of rabbits. Introduced plant effect is moderate- Carrichtera annua common t
hroughout the site. Homogeneity- community continues another 50m north of the road which is c
lose to the northern edge of the plot. 20m to the east and 50m to the west where there is a l
ow rise. 100m to the south along the drainage line.
## 2
Slope of low hill.
## 3
<NA>
## 4
On top of the James Range. Long unburnt. Weed impact minimal. Grazing impact nil.
## 5
On top of the James Range. Long unburnt. Weed impact minimal. Grazing impact nil.
## 6
2 km. south east of the Flux tower SASMDD0001, 100m in easterly direction from track.
##   outcrop_lithology other_outcrop_lithology plot_dimensions
## 1                LI                      LI     100 x 100 m.
## 2                NC                      NC     100 x 100 m.
## 3                NC                      NC     100 x 100 m.
## 4                SA                       M     100 x 100 m.
## 5                SA                       M     100 x 100 m.
## 6                NC                      NC     100 x 100 m.
##   site_location_visit_id    visit_start_date      visit_end_date
## 1                  56932 2014-09-07T15:24:00 2014-09-07T15:24:00
## 2                  53749 2012-10-25T00:00:00 2012-10-25T00:00:00
## 3                  53595 2013-04-26T00:00:00 2013-04-26T00:00:00
## 4                  58010 2016-03-02T00:00:00 2016-03-02T00:00:00
## 5                  53624 2011-10-06T00:00:00 2011-10-06T00:00:00
## 6                  53711 2012-09-23T00:00:00 2012-09-23T00:00:00
##
visit_notes
## 1
```

```
                                                              <NA>
## 2
## 3 Acacia aneura woodland. Rising a little more to the NE corner where it becomes rockier.
Site is on Plateau above Mitchell grass downs 10km to the north. Grazing impact low- lots of
echidna diggings in the site plus kangaroos. Weed impact low. Very long unburnt- no scarring
on Ironbark or Mulga and a varied cohort with old trees and younger shrubs.
## 4
Revisit collected Point intercept, vouchered plant specimens,DNA and Metagenomic samples.
## 5
                                                              <NA>
## 6
Low Mallee woodland in dune swale with dominant species of Eucalyptus oleosa subspecies oleos
a and Eucalyptus dumosa.
##                                                             location_description
## 1                                           Great Victoria Desert Nature Reserve
## 2     Umbeara Station 26.5km South East of Umbeara Homestead. 11km North of SA/Nt Border
## 3                                                 61km SE of Issiford on Idalia NP
## 4                                                                     Owen Springs
## 5 Top of James Range, Owen Springs Reserve approximately 73km South West of Alice Springs
## 6                                               Mallee swale. 23km North West of Renmark
##    erosion_type erosion_abundance erosion_state microrelief drainage_type
## 1          NC                 X            NC           Y              3
## 2           R                 1             A          NC              5
## 3          NC                 X            NC           N              4
## 4          NC                NC            NC          NC              7
## 5          NC                NC            NC          NC              2
## 6          NC                NC            NC          NC              7
##    disturbance climatic_condition vegetation_condition observer_veg
## 1           0                DRY                  FFR           18
## 2          1L                DRY                  FFR            1
## 3           0                DRY                  DRY            1
## 4          NC                DRY                  AVG            1
## 5           0                WET                  DRY            1
## 6          NC                DRY                  DRY            1
##    observer_soil described_by pit_marker_easting pit_marker_northing
## 1             2            1             383287             6676768
## 2             2            1             389476             7134938
## 3             2            1             875168             7257129
## 4            11            1                 NA                  NA
## 5             2            1             335077             7324080
## 6             2            1             462393             6236497
##    pit_marker_mga_zones pit_marker_datum pit_marker_location_method
## 1                    52            WGS84                        GPS
## 2                    53            WGS84                       DGPS
## 3                    55            WGS84                       DGPS
## 4                    53            WGS84                       <NA>
## 5                    53            GDA94                       DGPS
## 6                    54            GDA94                        GPS
##    soil_observation_type          a_s_c plot_is_100m_by_100m
## 1                     P           <NA>                  TRUE
## 2                     P  3RUCYCZARFLLT                  TRUE
## 3                     P           <NA>                  TRUE
## 4                    NC           <NA>                  TRUE
## 5                     P  3RUCYCZAIIKT                  TRUE
## 6                     P           <NA>                  TRUE
##    plot_is_aligned_to_grid plot_is_permanently_marked  latitude longitude
## 1                     TRUE                       TRUE -30.03548  127.7895
## 2                     TRUE                       TRUE -25.89989  133.8966
```

```
## 3                      TRUE                      TRUE -24.75512  144.7083
## 4                      TRUE                      TRUE -24.18724  133.3764
## 5                      TRUE                      TRUE -24.18724  133.3764
## 6                      TRUE                      TRUE -34.01170  140.5927
##   point state       site_unique
## 1    SW    WA WAANUL0007-56932
## 2    SW    NT NTAFIN0031-53749
## 3    SW   QLD QDAMUL0003-53595
## 4    SW    NT NTAFIN0004-58010
## 5    SW    NT NTAFIN0004-53624
## 6    SW    SA SASMDD0002-53711
```

```
# Explore 'veg_PI' data
dim(AP.data$veg.PI)
```

```
## [1] 734464      13
```

```
names(AP.data$veg.PI)
```

```
##  [1] "site_location_name"     "site_location_visit_id"
##  [3] "transect"               "point_number"
##  [5] "veg_barcode"            "herbarium_determination"
##  [7] "substrate"              "in_canopy_sky"
##  [9] "dead"                   "growth_form"
## [11] "height"                 "hits_unique"
## [13] "site_unique"
```

```
head(AP.data$veg.PI)
```

```
##   site_location_name site_location_visit_id transect point_number
## 1         WAAC000006                  53438    S1-N1            0
## 2         WAAC000006                  53438    S1-N1            1
## 3         WAAC000006                  53438    S1-N1            2
## 4         WAAC000006                  53438    S1-N1            3
## 5         WAAC000006                  53438    S1-N1            4
## 6         WAAC000006                  53438    S1-N1            5
##   veg_barcode            herbarium_determination substrate
## 1        <NA>                               <NA>    Crypto
## 2        <NA>                               <NA>    Crypto
## 3 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
## 4 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
## 5 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
## 6 WAA  001053 Melaleuca pauperiflora subsp. fastigiata    Litter
##   in_canopy_sky  dead growth_form height hits_unique       site_unique
## 1            NA    NA        <NA>     NA  S1-N1 0 WAAC000006-53438
## 2            NA    NA        <NA>     NA  S1-N1 1 WAAC000006-53438
## 3         FALSE FALSE       Shrub    3.5  S1-N1 2 WAAC000006-53438
## 4         FALSE FALSE       Shrub    3.5  S1-N1 3 WAAC000006-53438
## 5         FALSE FALSE       Shrub    3.7  S1-N1 4 WAAC000006-53438
## 6         FALSE FALSE       Shrub    3.5  S1-N1 5 WAAC000006-53438
```

MANIPULATING AusPlots DATA

The `get_ausplots` function extracts and compiles AusPlots data allowing substantial flexibility in the selection of the required data. Up to 8 different types of data can be retrieved into distinct data frames (i.e. data on sampling sites, vegetation structure, vegetation point intercept, vegetation vouchers, vegetation basal wedge, soil characterization, soil bulk density, and soil & soil metagenomics samples). In addition, data can be filtered for particular sets of plots and/or genus/species, as well as geographically using a rectangular bounding box.

However, in some situations we are only interested in a subset of the data retrieved by `get_ausplots`. To subset ausplot data, we use the variables in the retrieved data frames corresponding to the concept by we would like to filter the data. The retrieved data by the function 'get_ausplots' can be manipulated as any other R data. However, the 'deep' structure of the data (a list of multiple data frames) and interrelation of the data frames (via a common a common link variable) can make manipulating the data a bit more daunting.

An example, were we focus on the sites in the 5 most sampled Bioregions, is presented below. We will first identify which are these regions, and then subset the sites in these regions. Detailed explanations and examples of the archetypical scenarios encountered when sub-setting AusPlots data (retrieved with `ausplotsR`) are provided another TERN DSDP Tutorial "'AusplotsR' Package and AusPlots Data Basics" Tutorial (https://github.com/ternaustralia/TERN-Data-Skills/tree/master/EcosystemSurveillance_PlotData/AusPlots_BasicTutorial).

```
#-------------------------------------------------------------------------------
# Find the 5 most 'sampled' Bioregions
#-------------------------------------------------------------------------------

# Create a derived Bioregions Factor Variable in the 'site.info' DF
AP.data$site.info$bioregion.f = factor(AP.data$site.info$bioregion_name)
#names(AP.data$site.info)

# Display the Bioregions number of visits (from most visited to least visited)
sort(summary(AP.data$site.info$bioregion.f), decreasing=TRUE)
```

```
## MDD FLB SSD GFU STP PIL BHC MGD GUP COO RIV MAC CYP FIN AUA CHC NUL KAN
##  52  50  48  41  40  35  34  34  33  32  32  28  19  18  15  13  13  11
## SYB EIU MUL BRT HAM MUR STU GVD AVW SWA VIB ARP CEK DAC DMR EYB GAW GES
##   9   7   7   6   6   6   6   5   4   4   4   3   3   3   3   3   3   3
## JAF LSD MAL NSS PCK BBS COP GAS MII NAN DAB DAL ESP GSD
##   3   3   3   3   3   2   2   2   2   2   1   1   1   1
```

```
# Get the Names of the 5 most visited Bioregions
Bioregs.Top5.s = names(sort(summary(AP.data$site.info$bioregion.f), decreasing=TRUE)[1:5])
Bioregs.Top5.s
```

```
## [1] "MDD" "FLB" "SSD" "GFU" "STP"
```

```
#-------------------------------------------------------------------------------
# Subset data for the 5 most 'visited/sampled' Bioregions
#-------------------------------------------------------------------------------

summary(AP.data)
```

```
##           Length Class      Mode
## site.info 44     data.frame list
## veg.basal 10     data.frame list
## veg.PI    13     data.frame list
## citation  1      -none-     character
```

```
# Subset the 5 most sampled Bioregions in the 'site.info' data frame
# ==================================================================
dim(AP.data$site.info)
```

```
## [1] 662  44
```

```
AP.BioregTop5.l = AP.data
AP.BioregTop5.l$site.info = AP.BioregTop5.l$site.info[AP.BioregTop5.l$site.info$bioregion_nam
e %in% Bioregs.Top5.s, ]
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231  44
```

```
# Drop unused levels in the bioregion.f factor (i.e. the levels corresponding to other
# bioregions are dropped).
levels(AP.BioregTop5.l$site.info$bioregion.f)
```

```
##  [1] "ARP" "AUA" "AVW" "BBS" "BHC" "BRT" "CEK" "CHC" "COO" "COP" "CYP"
## [12] "DAB" "DAC" "DAL" "DMR" "EIU" "ESP" "EYB" "FIN" "FLB" "GAS" "GAW"
## [23] "GES" "GFU" "GSD" "GUP" "GVD" "HAM" "JAF" "KAN" "LSD" "MAC" "MAL"
## [34] "MDD" "MGD" "MII" "MUL" "MUR" "NAN" "NSS" "NUL" "PCK" "PIL" "RIV"
## [45] "SSD" "STP" "STU" "SWA" "SYB" "VIB"
```

```
AP.BioregTop5.l$site.info$bioregion.f = droplevels(AP.BioregTop5.l$site.info$bioregion.f)
levels(AP.BioregTop5.l$site.info$bioregion.f)
```

```
## [1] "FLB" "GFU" "MDD" "SSD" "STP"
```

```
# Subset the 5 most sampled Bioregions in the 'veg.PI' data frame
# ================================================================
# Because we are just subsetting the sites within the 5 most
# sampled bioregions, using the variable 'site_location_name'
# is enough (i.e. we don't need to use the variable 'site_unique').
dim(AP.BioregTop5.l$veg.PI)
```

```
## [1] 734464     13
```

```
AP.BioregTop5.l$veg.PI = AP.BioregTop5.l$veg.PI[AP.BioregTop5.l$veg.PI$site_location_name %in
% AP.BioregTop5.l$site.info$site_location_name, ]
dim(AP.BioregTop5.l$veg.PI)
```

```
## [1] 245319     13
```

```
# Subset the 5 most sampled Bioregions in the 'veg.basal' data frame
# ================================================================
# Because we are just subsetting the sites within the 5 most
# sampled bioregions, using the variable 'site_location_name'
# is enough (i.e. we don't need to use the variable 'site_unique').
dim(AP.BioregTop5.l$veg.basal)
```

```
## [1] 8291    10
```

```
AP.BioregTop5.l$veg.basal = AP.BioregTop5.l$veg.basal[AP.BioregTop5.l$veg.basal$site_location
_name %in% AP.BioregTop5.l$site.info$site_location_name, ]
dim(AP.BioregTop5.l$veg.basal)
```

```
## [1] 2323    10
```

MAP THE SITES

Next we visualise the sites on a map of Australia. First we graph all the Sites curently in AusPlots and then the Sites in the 5 most sampled bioregions. To do so we first obtain the map from the `maps` package and convert it to `SpatialPolygons`. Then we plot the Sites on the SpatialPolygon object for the map of Australia using functions in the `ggplot2` package. To differenciate among bioregions, sites are represented by different shapes and colours in the first graph, and by dots of different colours in the second one.

```
#-------------------------------------------------------------------------------
# Get and Prepare a Map of Australia
#-------------------------------------------------------------------------------

# Maps in the package 'maps' are projected in longlat by default
aus = map("worldHires", "Australia", fill=TRUE, xlim=c(110,160),ylim=c(-45,-5), mar=c(0,0,0,0
), plot=FALSE)

# Convert map data to SpatialPolygons
#aus.sp = map2SpatialPolygons(aus, IDs=aus$names, proj4string=CRS("+proj=longlat"))
CRS("+init=epsg:4326") # More info (i.e. provides a datum)
```

```
## CRS arguments:
##  +init=epsg:4326 +proj=longlat +datum=WGS84 +no_defs +ellps=WGS84
## +towgs84=0,0,0
```

```r
aus.sp = map2SpatialPolygons(aus, IDs=aus$names, proj4string=CRS("+init=epsg:4326"))


#--------------------------------------------------------------------------------
# Plot All AusPlots Sites on a Map of Australia
#--------------------------------------------------------------------------------

# We well use 7 distinct symbol shapes to represent the AusPlots Sites (combined with
# different colors). We will cycle through the 7 symbol shapes. We start by creating a
# vector of with symbol shapes codes as long as the number of bioregions in the current
# version of the AusPlots dataset, cycling among the 7 shapes. What complicates this
# process a bit is that the number of bioregions sampled changes with time, as additional
# sites in different bioregions are sampled. Thus, we need to estimate the required number
# of cycles of symbol shapes from the data. To do this we use the function ceiling, and
# the cut back to the required number of symbol shapes as we might not need full cycles
# (i.e. the number of sites might not be a multiple of 7; e.g. 50 sites require more than
# 7 cycles, so we use 8 and then trimm the vector from 56 (7 shapes * 8 cycles) to 50.

# Preparation: Create a vector with the symbol shapes values
bioregions.cnt = length(levels(AP.data$site.info$bioregion.f))
shape.cycles.num = ceiling(bioregions.cnt / 7) #  Using 7 distinct Symbol Shapes
sites.shape.values = rep(c(15:18,3:4,8),shape.cycles.num)[1:bioregions.cnt]

# Create Plot
ggplot( data=AP.data$site.info,
               aes(x = longitude, y = latitude, group=bioregion.f), alpha =0.5) +
geom_point(aes(colour=bioregion.f, fill=bioregion.f, shape=bioregion.f), size=1.5) +
scale_shape_manual(values=sites.shape.values) +  # Cycle through Symbol Types
ggtitle("Locations of all AusPlots sites") +
theme(plot.title = element_text(hjust = 0.5, face="bold", size=14)) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)
```

Locations of all AusPlots sites
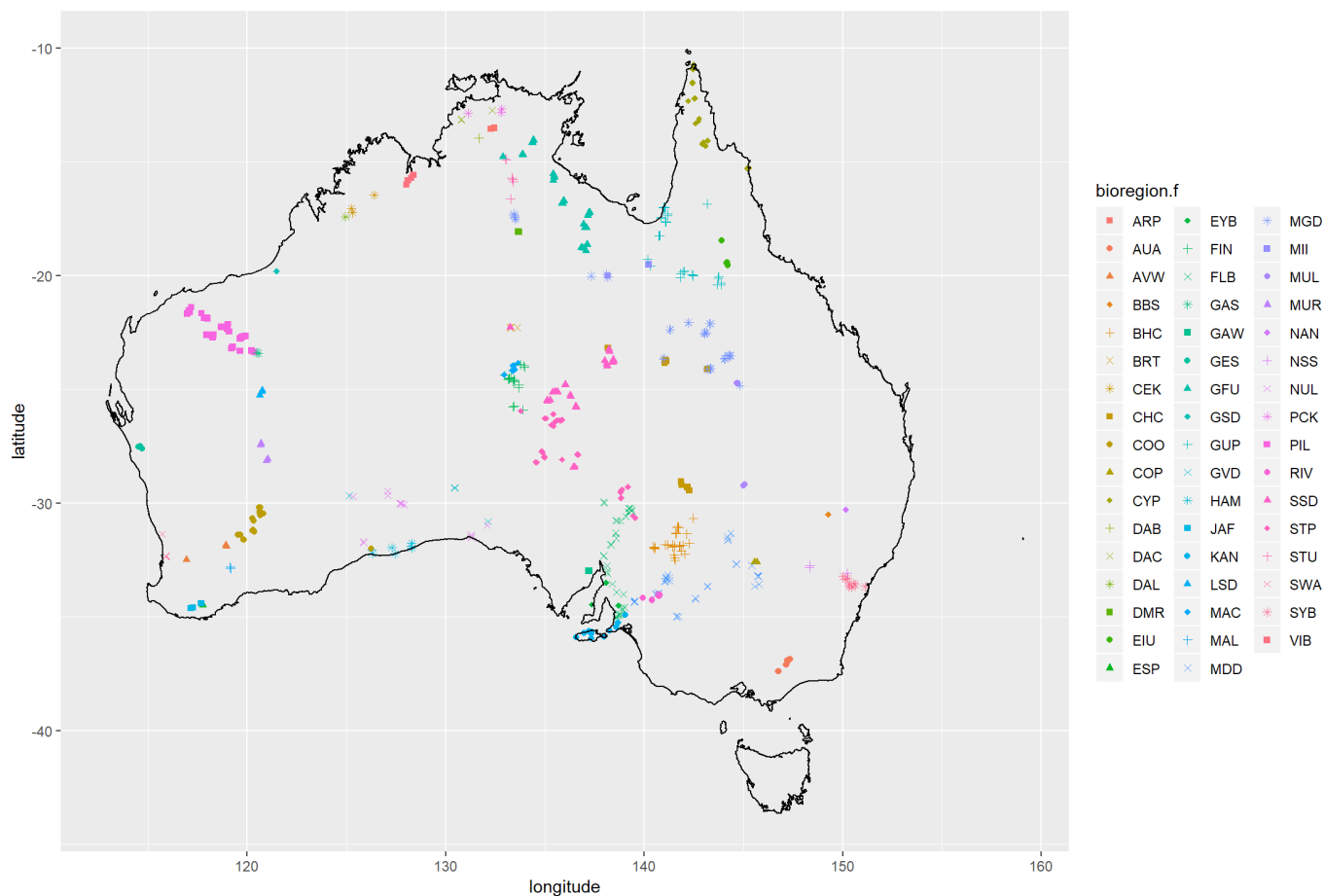

```
#-----------------------------------------------------------------------------
# Plot AusPlots sites in the 5 Bioregions on Map of Australia
#-----------------------------------------------------------------------------
ggplot(data=AP.BioregTop5.l$site.info, aes(x = longitude, y = latitude, colour=bioregion.f, f
ill=bioregion.f), alpha =0.5) +
geom_point(pch=21, size=1.5) + scale_colour_manual(values = c("darkgreen", "magenta", "red",
"yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
ggtitle("Locations of sites in the 5 most sampled bioregions") +
theme(plot.title = element_text(hjust = 0.5, face="bold", size=14)) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)
```

Locations of sites in the 5 most sampled bioregions
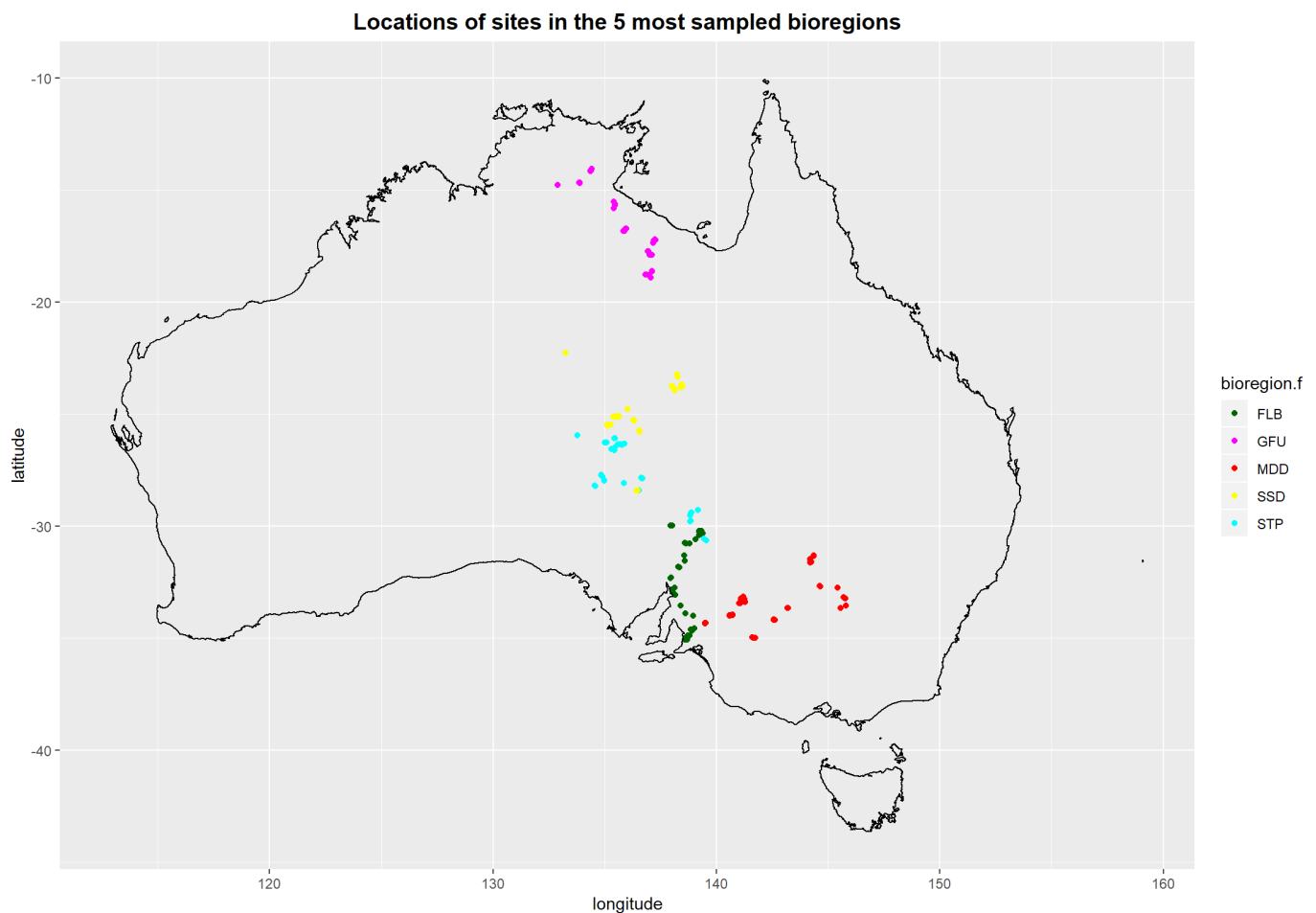
SPECIES-LEVEL DATA: `species_table` function and species occurrence matrices

In this section, we will explore to how to obtain and use species occurrence data from AusPlots raw data. In particular, we will examine species cover/abundance, species presence/absence, multiple indices of species diversity, and rank-abundance plots for the sites in the 5 most sampled bioregions.

First step: Create a species occurrence matrix

The first step to work with species-level AusPlots data is to create a species occurrence matrix. The `species_table` function in the `ausplotsR` package can be used to effortlessly create this type of matrix. This function takes a data frame of individual raw point intercept hits (i.e. a `veg.PI` data frame) generated using the `get_ausplots` function and returns a 'species against sites' matrix. Four metrics can be selected to score species occurrence:

- *Presence/Absence*: Set by the argument `m_kind = PA`.
- *Percent Cover*: Based on total frequency of hits. This is the most commonly used metric. Set by the argument `m_kind = percent_cover`.
- *Frequency*: Based on proportional frequencies of presence on the 10 individual transects within a plot. Set by the argument `m_kind = freq`. It can be a measure of importance for low cover species.
- *IVI*: A combination of cover and frequency. Set by the argument `m_kind = IVI`.

If Percent Cover or IVI are used two types of cover type can be selected:

- *Projected Foliage Cover* (*PFC*): Hits scored as 'in canopy sky' are removed. Set by the argument `cover_type = PFC`.
- *Opaque Canopy Cover* (*OCC*): Hits scored as 'in canopy sky' are retained. Set by the argument `cover_type = OCC`.

```r
# Use function 'species_table' in 'ausplotsR' package to create an Abundance per Site Table
# ==============================================================================
SppBYSites.BioregTop5 = species_table(AP.BioregTop5.l$veg.PI, m_kind="percent_cover", cover_t
ype="PFC")
class(SppBYSites.BioregTop5)
```

```
## [1] "data.frame"
```

```r
dim(SppBYSites.BioregTop5) # Number of rows and columns in the matrix: Sites x Species
```

```
## [1]  225 1220
```

```r
SppBYSites.BioregTop5[1:5, 1:5]
```

```
##                 Abutilon.fraseri Abutilon.halophilum Abutilon.otocarpum
## NSAMDD0001-56965                0                   0                  0
## NSAMDD0002-56952                0                   0                  0
## NSAMDD0003-56968                0                   0                  0
## NSAMDD0004-56953                0                   0                  0
## NSAMDD0005-56969                0                   0                  0
##                 Abutilon.sp. Acacia.adoxa
## NSAMDD0001-56965            0            0
## NSAMDD0002-56952            0            0
## NSAMDD0003-56968            0            0
## NSAMDD0004-56953            0            0
## NSAMDD0005-56969            0            0
```

```r
# Enrich Table with: Site_Location, Bioregion, Latitude, and Longitude
# =====================================================================

# Create a 'site_unique' variable in Species by Sites Table to relate both datasets
# --------------------------------------------------------------------------------
SppBYSites.BioregTop5$site_unique = rownames(SppBYSites.BioregTop5)

# Both DF have differente number of rows!
dim(SppBYSites.BioregTop5)
```

```
## [1]  225 1221
```

```r
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231   44
```

```
# Enrich with: Bioregion, Latitude, and Longitude
# -------------------------------------------------
SppBYSites.BioregTop5  = merge(SppBYSites.BioregTop5, AP.BioregTop5.l$site.info,
by="site_unique")[,c(names(SppBYSites.BioregTop5),
                                    "bioregion.f", "longitude", "latitude")]
SppBYSites.BioregTop5 = na.omit(SppBYSites.BioregTop5)
#head(SppBYSites.BioregTop5)
#summary(SppBYSites.BioregTop5)
head(names(SppBYSites.BioregTop5))
```

```
## [1] "Abutilon.fraseri"    "Abutilon.halophilum" "Abutilon.otocarpum"
## [4] "Abutilon.sp."        "Acacia.adoxa"        "Acacia.alleniana"
```

Species Abundance

In AusPlots data, vegetation 'percent cover' is used as a surrogate of vegetation 'abundance'. In this section, we will examine percent cover by:

- Site visit and species: That is, all cells in the 'Species by Sites' table.
- Species: By computing the column totals in the 'Species by Sites' table.

Percent Cover (Abundance) by Site Visit x Species

```
# Minimum and Maximum Site Visit x Species Abundance values
# -----------------------------------------------------------
# '-4' because we added 4 new columns (Plot, bioregion.f, longitude, and latitude)##
range(SppBYSites.BioregTop5[,1: (dim(SppBYSites.BioregTop5)[2]-4)])
```

```
## [1]  0.00000 81.18812
```

```
# Plot Highest  Site Visit x Species Abundance values
# -----------------------------------------------------
Abundance = unlist(SppBYSites.BioregTop5[,1: (dim(SppBYSites.BioregTop5)[2]-4)])
head(Abundance)
```

```
## Abutilon.fraseri1 Abutilon.fraseri2 Abutilon.fraseri3 Abutilon.fraseri4
##                 0                 0                 0                 0
## Abutilon.fraseri5 Abutilon.fraseri6
##                 0                 0
```

```
#length(Abundance)
#dim(SppBYSites.BioregTop5)

# Plot the 5 Site Visits x Species combination with the Highest Abundances
par(mfrow=c(1,1))
barplot(sort(Abundance, decreasing=TRUE)[1:5], las=1, xlab="Abundance class", ylab="Frequenc
y")
```
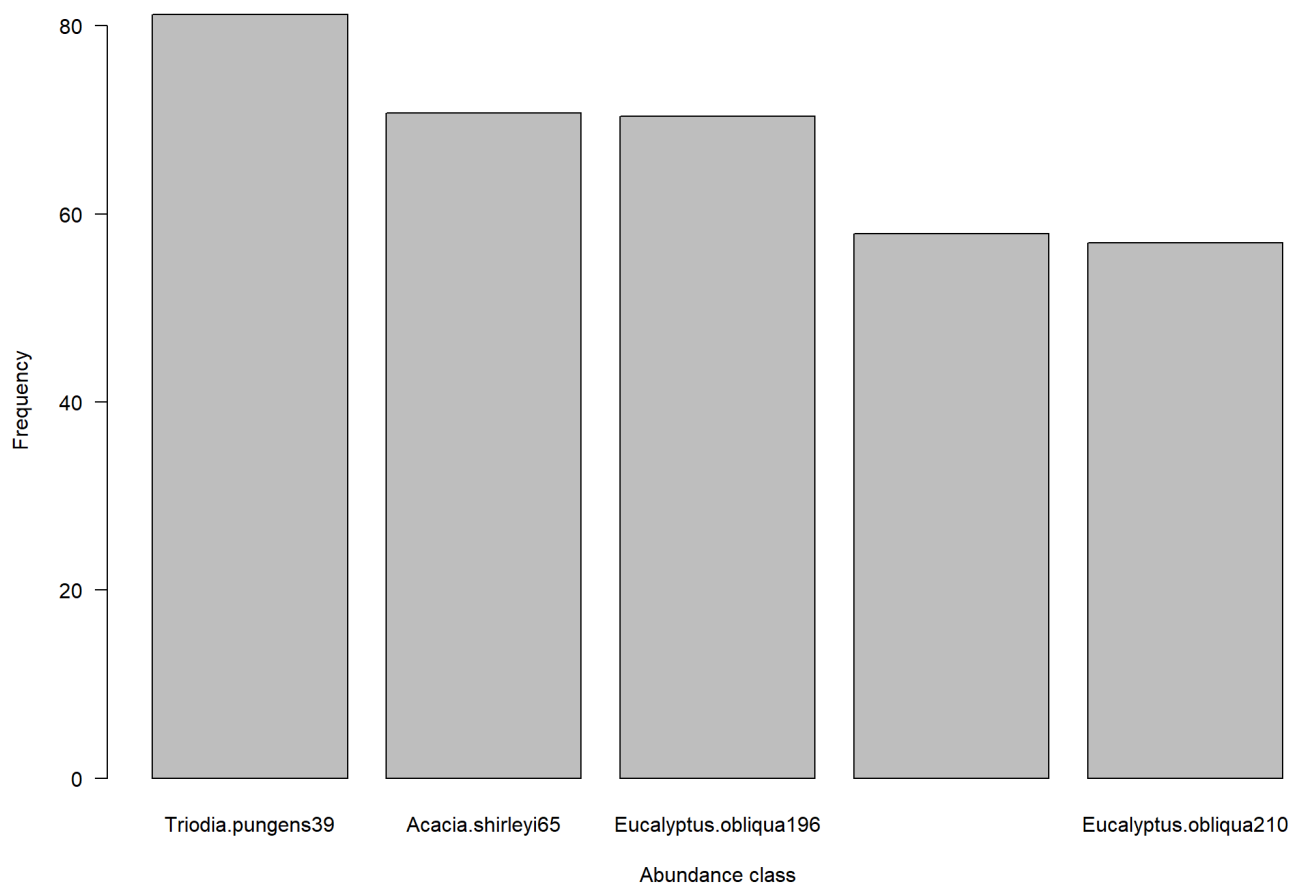
```
# ggplot2 graph to make it look nicer. Now we plot the 25 species-site visit covers with hori
zontal bars
temp.labs = names(sort(Abundance, decreasing=TRUE))
temp.Abundances = sort(Abundance, decreasing=TRUE)
temp.df = data.frame(temp.Abundances, temp.labs)
# Order factor levels so that bars are sorted by Abundance in the plot. Otherwise they would
 plotted in alphabetical order
# 'rev' to plot bars in decreasing order (i.e. larger bar at top; otherwise larger bar at bot
tom)
temp.df$temp.labs = factor(temp.df$temp.labs, levels=rev(temp.df$temp.labs))
ggplot(data=temp.df[1:25,], aes(x=temp.labs, y=temp.Abundances)) +
geom_bar(stat="identity", fill="steelblue") +
geom_text(aes(label=round(temp.Abundances,2)), hjust=-0.1, size=4)+
labs(x="Species & Site Visit", y="Abundance (Cover %)") +
theme_minimal() + coord_flip()
```
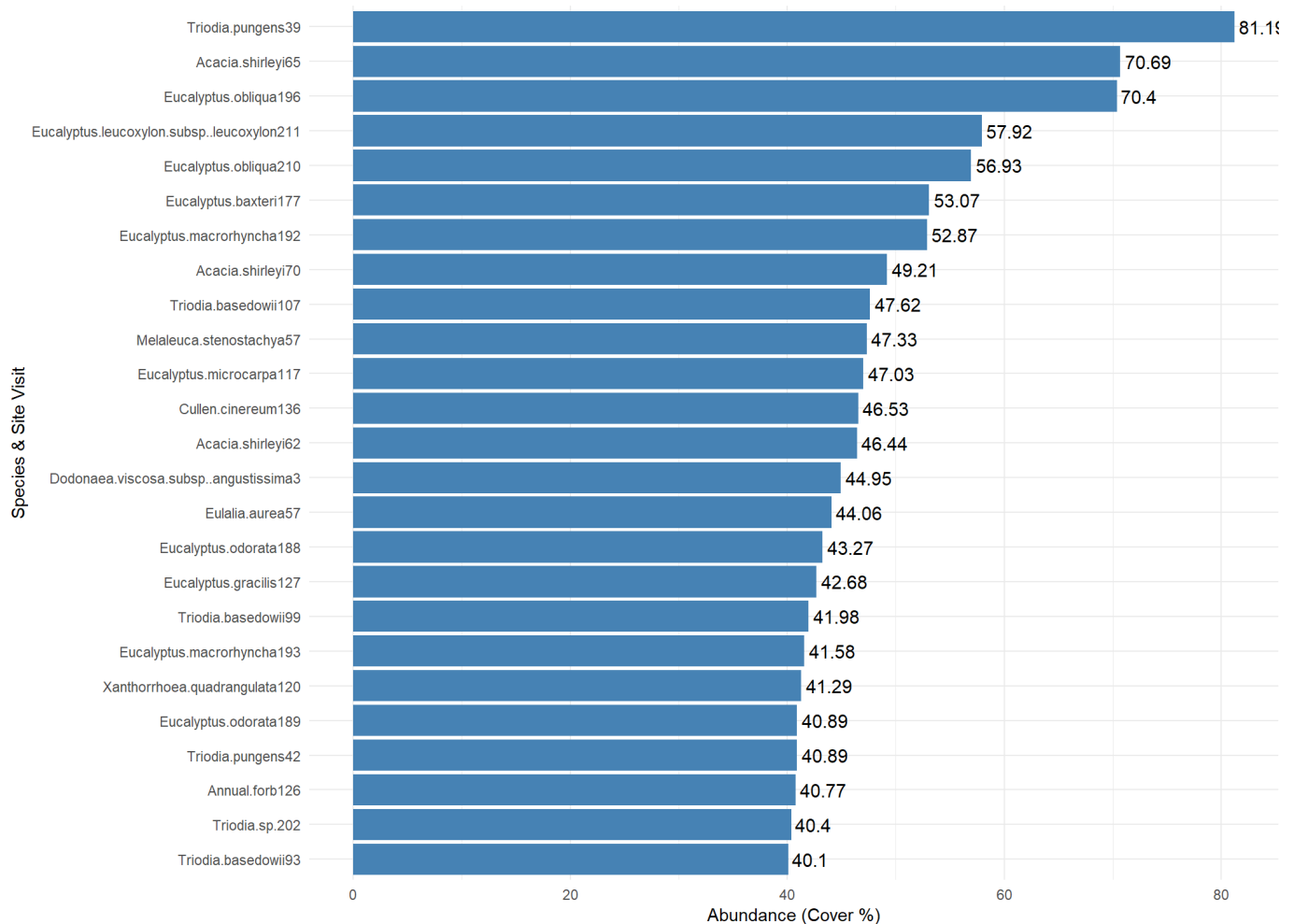
```
# Cleaning up
rm(list=ls(pattern="temp."))
```

Abundance (Cover %) by Species

Now we compute the percent cover of all species across the sites in the 5 most sampled bioregions. Then we find and plot on a map of Australia the 4 most Abundant species in the 5 regions (across all regions pooled together).

```
# Compute Species Total Abundance (Cover %)
# -----------------------------------------
TotAbundances.BioregTop5 = colSums(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4
)])
head(TotAbundances.BioregTop5)
```

```
##      Abutilon.fraseri Abutilon.halophilum  Abutilon.otocarpum
##            0.1980198           1.3861386           2.2787055
##         Abutilon.sp.        Acacia.adoxa     Acacia.alleniana
##            0.5944519           0.4950495           1.1881188
```

```
# Species with Highest Total Abundance
# ------------------------------------
# Species with Highest Total Abundance
max(TotAbundances.BioregTop5)
```

```
## [1] 321.4717
```

```
which.max(TotAbundances.BioregTop5)
```

```
## Triodia.basedowii
##               1155
```

```
# Species with Top 4 Highest Abundances
TotAbundances4Highest.indices =
    which(TotAbundances.BioregTop5 >= sort(TotAbundances.BioregTop5, decreasing=T)[4],
arr.ind=T)
sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices], decreasing=TRUE)
```

```
##     Triodia.basedowii      Triodia.bitextura Schizachyrium.fragile
##              321.4717               283.4677              256.2688
##       Triodia.pungens
##              251.6832
```

```r
# Plot 4 Species with Highest Cover in the 5 Most Sampled Bioregions
# -----------------------------------------------------------------

# Most Abundant Species
spp = names(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE))[1]
plot.title =  paste(spp, " (total cover over the 5 bioregions = ",
                    round(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE)[1],2) , ")", sep="")
TotAbundance.spp1 =
ggplot(data=SppBYSites.BioregTop5, aes(x=longitude, y=latitude, colour=bioregion.f, fill=bior
egion.f), alpha =0.5) +
geom_point(aes_string(size=spp), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
ggtitle(plot.title) + theme(plot.title = element_text(hjust = 0.5)) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# 2nd Most Abundant Species
spp = names(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE))[2]
plot.title =  paste(spp, " (total cover over the 5 bioregions = ",
                    round(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE)[2],2) , ")", sep="")
TotAbundance.spp2 =
ggplot(data=SppBYSites.BioregTop5, aes(x=longitude, y=latitude, colour=bioregion.f, fill=bior
egion.f), alpha =0.5) +
geom_point(aes_string(size=spp), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
ggtitle(plot.title) + theme(plot.title = element_text(hjust = 0.5)) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# 3rd Most Abundant Species
spp = names(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE))[3]
plot.title =  paste(spp, " (total cover over the 5 bioregions = ",
                    round(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE)[3],2) , ")", sep="")
TotAbundance.spp3 =
ggplot(data=SppBYSites.BioregTop5, aes(x=longitude, y=latitude, colour=bioregion.f, fill=bior
egion.f), alpha =0.5) +
geom_point(aes_string(size=spp), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
ggtitle(plot.title) + theme(plot.title = element_text(hjust = 0.5)) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# 4th Most Abundant Species
spp = names(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE))[4]
plot.title =  paste(spp, " (total cover over the 5 bioregions = ",
                    round(sort(TotAbundances.BioregTop5[TotAbundances4Highest.indices],
decreasing=TRUE)[4],2) , ")", sep="")
TotAbundance.spp4 =
ggplot(data=SppBYSites.BioregTop5, aes(x=longitude, y=latitude, colour=bioregion.f, fill=bior
egion.f), alpha =0.5) +
```

```
geom_point(aes_string(size=spp), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
ggtitle(plot.title) + theme(plot.title = element_text(hjust = 0.5)) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# Plot the 4 Graphs
grid.arrange(TotAbundance.spp1, TotAbundance.spp2, TotAbundance.spp3, TotAbundance.spp4, nrow
=2)
```
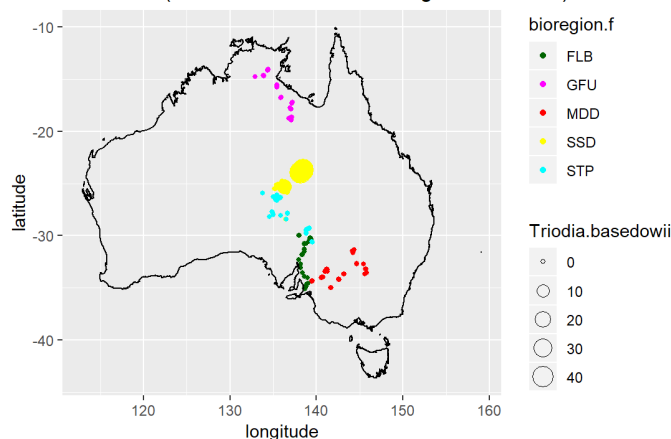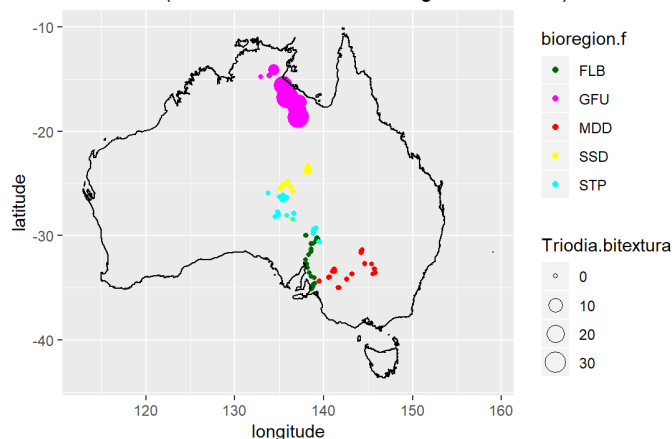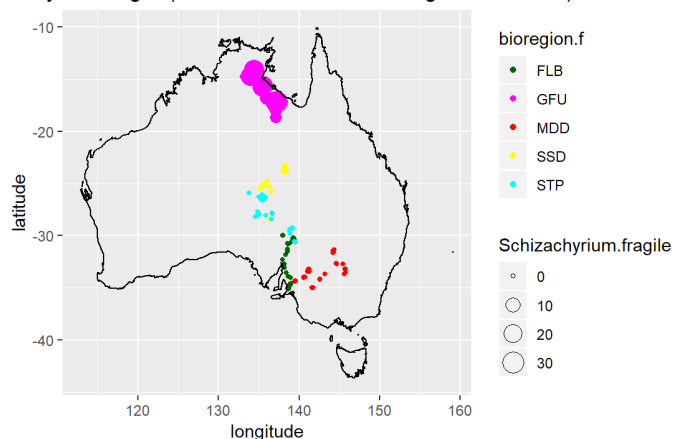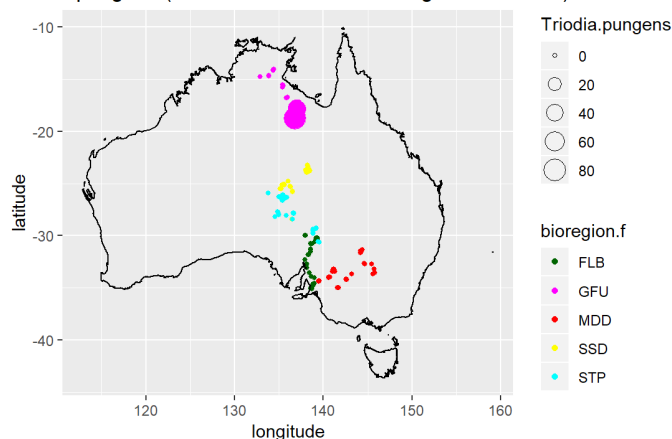

Species Occurrence (Presence/Absence)

We next focus on species occurrence data; that is, whether as species is Present/Absent. We can compute Presence/Absence data in several ways:

- Re-use the 'percent cover' data we used above: If percent cover is > 0 then the species is present, if percent cover = 0 then the species is absent.
- Use the `species_table` function with the argument `m_kind = PA` on the data frame with raw point intercept data (i.e. `veg.PI`) generated by the function `get_auplots` .
- Use the `species_table` function with the argument `m_kind = PA` on the data frame with vegetation vouchers data (a complete set of species records for the plot determined by a herbarium; i.e. `veg.vouchers`) generated by the function `get_auplots` . This option provides the most complete species inventories by sites.

Here we use the first option for simplicity and generality (i.e. how these tasks can be performed with other abundance data outside AusPlots). As for the Abundance/Percent Cover data, we first examine species occurrence across all cells (i.e. combinations of site visits and species), and then investigate and plot the total number of (absolute and relative) occurrences for each species.

```
# Presence/Absence across all Cells (i.e. Site Visit x Species)
# ============================================================

# Absolute Presences/Absences
# ---------------------------
 # Number of Presences
sum(Abundance > 0)
```

```
## [1] 4394
```

```
 # Number of Absences
sum(Abundance == 0)
```

```
## [1] 270106
```

```
# Relative (%) Presences/Absences
# -------------------------------
num.cells = (nrow(SppBYSites.BioregTop5[1: (dim(SppBYSites.BioregTop5)[1]-3)]) * ncol(SppBYSi
tes.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)]))
 # % of Presences
sum(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)] > 0)/ num.cells
```

```
## [1] 0.01600729
```

```
 # % of Absences
sum(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)] == 0) / num.cells
```

```
## [1] 0.9839927
```

```
# Total Presence/Absence  for each Species (i.e. per data frame Column)
# =====================================================================

# Compute number of sites where each species is present (sum by columns)
head(names(SppBYSites.BioregTop5)) # Species are in columns
```

```
## [1] "Abutilon.fraseri"   "Abutilon.halophilum" "Abutilon.otocarpum"
## [4] "Abutilon.sp."        "Acacia.adoxa"        "Acacia.alleniana"
```

```
SppPres.BioregTop5 = apply(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)]>0, 2,s
um)

# Sort results in increasing order
head(sort(SppPres.BioregTop5, decreasing = TRUE))
```

```
##           Annual.forb         Annual.grass       Dead.tree.shrub
##                   106                   89                    49
##       Sida.fibulifera    Aristida.holathera Schizachyrium.fragile
##                    39                   36                    31
```

```
# Compute Percentage Frequencies
SppRelFreq.BioregTop5 =  SppPres.BioregTop5 * 100 / nrow(SppBYSites.BioregTop5)

# Plot Species Frequencies
par(mfrow=c(1,1))
spp.freq.p1 = ggplot() + geom_histogram(aes(SppPres.BioregTop5)) +
              ggtitle("Species Sites Occurrence") +
              theme(plot.title = element_text(hjust = 0.5)) +
              xlab("Number of occurrences") + ylab("Number of species")
spp.freq.p2 = ggplot() + geom_histogram(aes(SppRelFreq.BioregTop5)) +
              ggtitle("Species Relative Frequencies") +
              theme(plot.title = element_text(hjust = 0.5)) +
              xlab("Frequency of occurrences (%)") + ylab("Number of species")
grid.arrange(spp.freq.p1, spp.freq.p2, ncol=2)
```
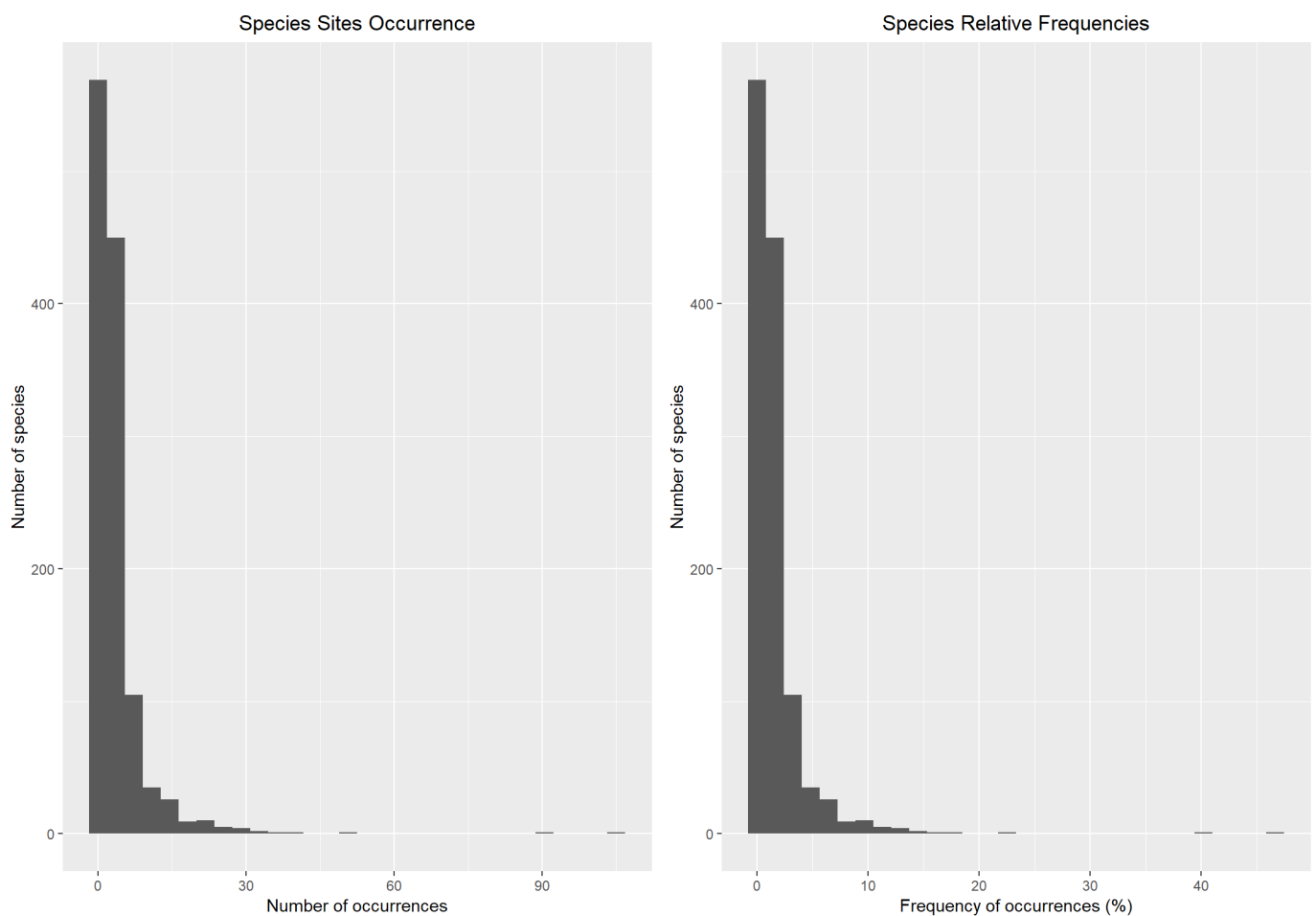

Species Diversity

On our exploration of the use of Species-lelvel AusPlots data, we now focus on Species Diversity. We first compute 7 common diversity indices, which we then place in a dataset. Finally, as an example, we plot two of these indices (Species Richness and Shanon Diversity Index) for the sites in the 5 most sampled bioregions on a map of Australia.

NOTE: Diversity indices were originally designed to be used with counts of number of individuals per species, rather than percent cover, as a measure of abundance. These indices are also used with percent cover in the literature (see Tomasckik and Sander, 1987 for an example using coral cover). We need, however, to be aware of the different kind of answers and interpretation of the results required.

```r
# Compute and place in a DF the Species Diversity Indices
# ========================================================

# Species Richness
N0 = rowSums(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)] > 0)
# Shannon Entropy
H = diversity(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)])
# Shannon Diversity Index
N1 = exp(H)
# Simpson Diversity Index
N2 = diversity(SppBYSites.BioregTop5[,1:(dim(SppBYSites.BioregTop5)[2]-4)], "inv")
# Shannon Evenness (Hill's ratio)
E1 = N1/N0
# Simpson Evenness (Hill's ratio)
E2 = N2/N0
# Pielou Evenness
J = H/log(N0)

# Create a Data Frame with the Species Diversity Indices
SppBYSites.BioregTop.Div.df = data.frame(N0, H, N1, N2, E1, E2, J)



# Map Species Richness and Shannon Diversity Index (as an example)
# ================================================================

# Add extra info to DataFrame (Bioregions, Longitude, and latitude)
SppBYSites.BioregTop.Div.df$bioregion.f = SppBYSites.BioregTop5[rownames(SppBYSites.BioregTop.Div.df),"bioregion.f"]
SppBYSites.BioregTop.Div.df$longitude = SppBYSites.BioregTop5[rownames(SppBYSites.BioregTop.Div.df),"longitude"]
SppBYSites.BioregTop.Div.df$latitude = SppBYSites.BioregTop5[rownames(SppBYSites.BioregTop.Div.df),"latitude"]
summary(SppBYSites.BioregTop.Div.df)
```

```
##        N0              H               N1              N2
## Min.   : 1.00   Min.   :0.000   Min.   : 1.000   Min.   : 1.000
## 1st Qu.:14.00   1st Qu.:1.410   1st Qu.: 4.096   1st Qu.: 2.830
## Median :19.00   Median :1.809   Median : 6.103   Median : 4.246
## Mean   :19.53   Mean   :1.809   Mean   : 7.084   Mean   : 4.754
## 3rd Qu.:25.00   3rd Qu.:2.220   3rd Qu.: 9.209   3rd Qu.: 6.202
## Max.   :43.00   Max.   :3.103   Max.   :22.257   Max.   :15.240
##
##        E1               E2                J            bioregion.f
## Min.   :0.08624   Min.   :0.05508   Min.   :0.1727   FLB:48
## 1st Qu.:0.26863   1st Qu.:0.17313   1st Qu.:0.5511   GFU:41
## Median :0.35648   Median :0.23439   Median :0.6468   MDD:50
## Mean   :0.37763   Mean   :0.26445   Mean   :0.6277   SSD:46
## 3rd Qu.:0.45174   3rd Qu.:0.31907   3rd Qu.:0.7215   STP:40
## Max.   :1.00000   Max.   :1.00000   Max.   :0.9610
##                                     NA's   :1
##    longitude         latitude
## Min.   :132.9   Min.   :-35.08
## 1st Qu.:136.3   1st Qu.:-33.09
## Median :138.2   Median :-28.43
## Mean   :138.3   Mean   :-27.35
## 3rd Qu.:139.3   3rd Qu.:-23.83
## Max.   :145.8   Max.   :-14.05
##
```

```r
# Create Species Richness Plot
Div.SR =
ggplot(data=SppBYSites.BioregTop.Div.df, aes(x=longitude, y=latitude, colour=bioregion.f, fil
l=bioregion.f), alpha =0.5) +
geom_point(aes_string(size=N0), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# Create Shanon Diversity Index Plot
Div.ShannonIndex =
ggplot(data=SppBYSites.BioregTop.Div.df, aes(x=longitude, y=latitude, colour=bioregion.f, fil
l=bioregion.f), alpha =0.5) +
geom_point(aes_string(size=N1), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# Plot the 2 graphs
grid.arrange(Div.SR, Div.ShannonIndex, ncol=2)
```
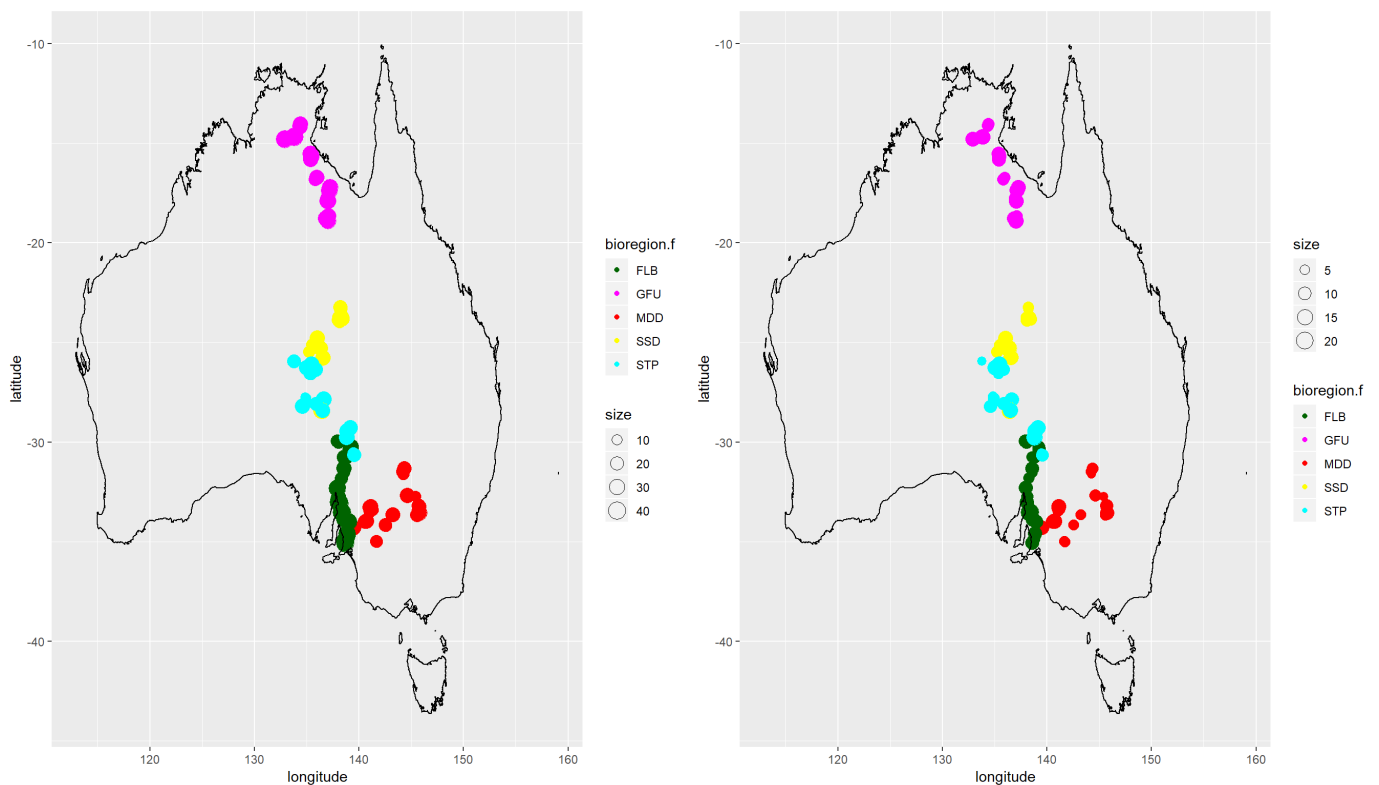
Rank-Abundance Curves & Relative Abundance Models

For the final example of downstream visualisation and analysis of Species-level AusPlots data, we focus on Rank-Abundance Curves (also known as Whittaker Plots). Rank-Abundance Curves provide further information on species diversity. They provide a more complete picture than a single diversity index. Their x-axis represents the abundance rank (from most to least abundant) and in the y-axis the species relative abundance. Thus, they depict both Species Richness (number of different in ranked) and Species Evenness (slope of the line that fits the rank; steep gradient indicates low evenness and a shallow gradient high evenness).

In this section we:

- We plot the Rank-Abundance Curves for the first 5 Site-Visits in our 5 most sampled bioregions dataset.
- We show a quick example of the fitting of possible Models of Relative Abundance for one Community.
- We compute the mean cover for each species in the 5 most sampled Bioregions and then plot the Rank-Abundance Curves for the 5 Bioregions.

```
par(mfrow=c(1,2))

# Rank-Abundance Curves (= Whittaker Plots) for the First 5 Site-Visits
# ===================================================================
goeveg::racurves(SppBYSites.BioregTop5[1:5,1:(dim(SppBYSites.BioregTop5)[2]-4)], bw=F)


# Possible Models of Relative Abundance for one Community
# ======================================================
plot(vegan::radfit(round(SppBYSites.BioregTop5[1,1:(dim(SppBYSites.BioregTop5)[2]-4)], digits
=0), log="xy"),
     pch=20, main="Rel. Abundance Models for Community NSAMDD0001-56965")
```

Rank-abundance diagram
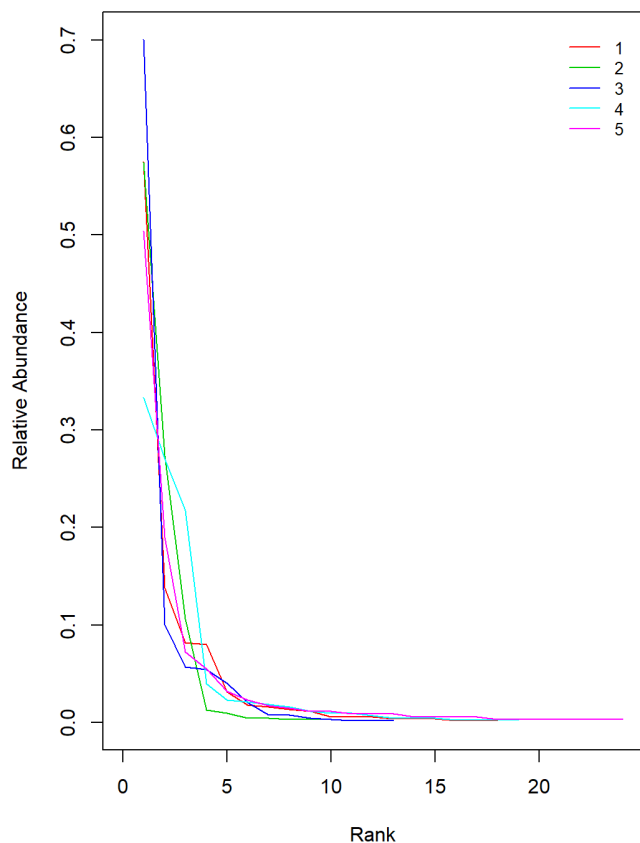
Rel. Abundance Models for Community NSAMDD0001-5696
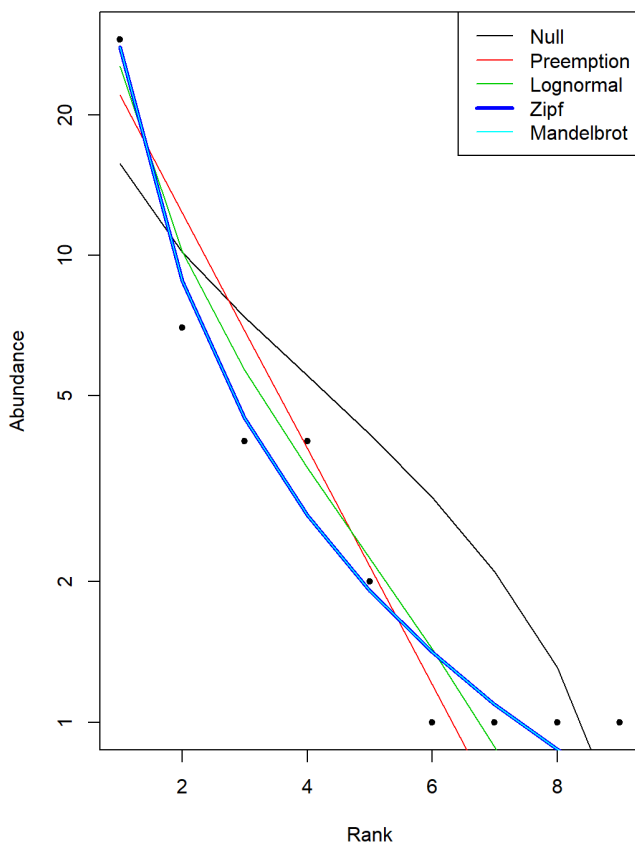
```
# Rank-Abundance Curves (= Whittaker Plots) for each Bioregion (using the Spp. Mean Cover)
# =============================================================================

# Compute Species Mean Cover for each of the 5 most sampled Bioregions
levels(SppBYSites.BioregTop5$bioregion.f)
```

```
## [1] "FLB" "GFU" "MDD" "SSD" "STP"
```

```
 # GFU
SppCover.GFU.Mean = apply(SppBYSites.BioregTop5[SppBYSites.BioregTop5$bioregion.f=="GFU",1:(d
im(SppBYSites.BioregTop5)[2]-4)], 2, mean)
#head(SppCover.GFU.Mean)
 # MDD
SppCover.MDD.Mean = apply(SppBYSites.BioregTop5[SppBYSites.BioregTop5$bioregion.f=="MDD",1:(d
im(SppBYSites.BioregTop5)[2]-4)], 2, mean)
#head(SppCover.MDD.Mean)
 # PIL
SppCover.PIL.Mean = apply(SppBYSites.BioregTop5[SppBYSites.BioregTop5$bioregion.f=="PIL",1:(d
im(SppBYSites.BioregTop5)[2]-4)], 2, mean)
#head(SppCover.PIL.Mean)
 # SSD
SppCover.SSD.Mean = apply(SppBYSites.BioregTop5[SppBYSites.BioregTop5$bioregion.f=="SSD",1:(d
im(SppBYSites.BioregTop5)[2]-4)], 2, mean)
#head(SppCover.SSD.Mean)
 # STP
SppCover.STP.Mean = apply(SppBYSites.BioregTop5[SppBYSites.BioregTop5$bioregion.f=="STP",1:(d
im(SppBYSites.BioregTop5)[2]-4)], 2, mean)
#head(SppCover.STP.Mean)

# Create a Matrix with Species Means per Bioregion
SppCover.BioregionMean.m = rbind(SppCover.GFU.Mean, SppCover.MDD.Mean, SppCover.PIL.Mean, Spp
Cover.SSD.Mean, SppCover.STP.Mean)
rownames(SppCover.BioregionMean.m)
```

```
## [1] "SppCover.GFU.Mean" "SppCover.MDD.Mean" "SppCover.PIL.Mean"
## [4] "SppCover.SSD.Mean" "SppCover.STP.Mean"
```

```
# Rank-Abundance Curves (= Whittaker Plots) for the Species Cover Mean in each of the 5 Biore
gions
par(mfrow=c(1,1))
goeveg::racurves(SppCover.BioregionMean.m, bw=F)
```

Rank-abundance diagram

Legend:
- SppCover.GFU.Mean
- SppCover.MDD.Mean
- SppCover.PIL.Mean
- SppCover.SSD.Mean
- SppCover.STP.Mean

(y-axis: Relative Abundance; x-axis: Rank)
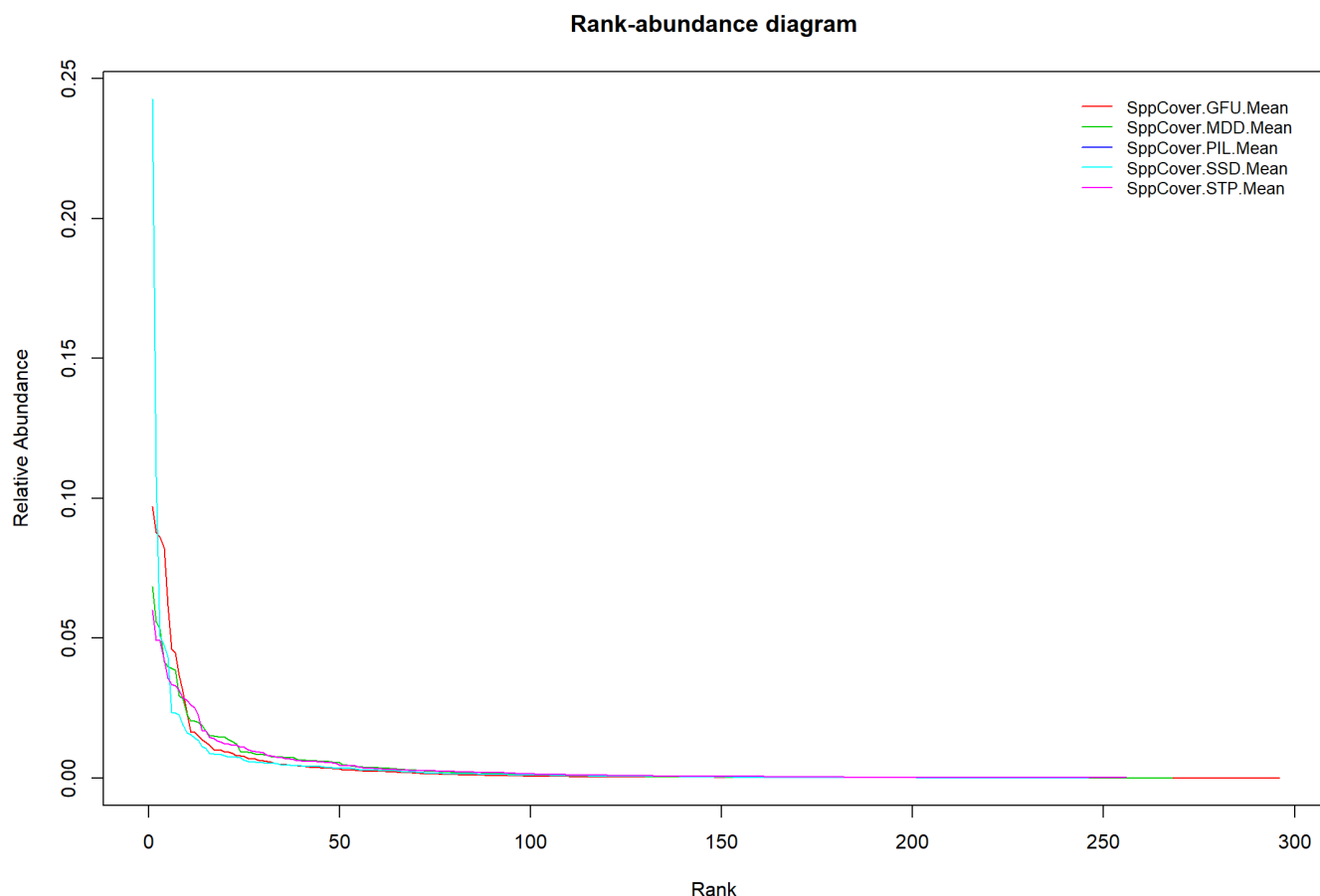
PROPORTIONAL VEGETATION COVER (= FRACTIONAL COVER): `fractional_cover` function

The `fractional_cover` function in the `auscoverR` package calculates fractional cover from AusPlots point intercept (PI) data (as generated by `get_ausplots`). Fractional cover refers to the proportions of green vegetation, dead vegetation and bare substrate cover. Cover fractions are assigned as follows:

- *'Green'* or *'photosynthetic vegetation'*: is living vascular plant cover.
- *'Brown'* or *'non-photosynthetic vegetation'*: is either vascular plant cover scored as 'dead', or substrate scored as litter, coarse woody debris or cryptogam (see below) that has no other veg cover.
- *'Bare'* or *'bare ground'*: is rock, outcrop, gravel or bare soil with no veg cover substrate.

Typically, 1010 intercept points are used to compute fractional cover in a plot. These 1010 intercept points correspond to 101 points per transect (i.e. 101 points at 1 meters intervals in each of the 100 metres long transect) x 10 transects (5 transects oriented North to South and 5 oriented East to West). Further details about the AusPlots methods can be found in TERN's DSDP AusPlots Methods Knowledge Base (https://ternaus.atlassian.net/wiki/spaces/ESM/pages/651329908/Ecosystem+Surveillance+Monitoring+-+AusPlots+Methods) and the AusPlots Rangelands Survey Protocol Manual (https://www.tern.org.au/AusPlots-Rangelands-Survey-Protocols-Manual-pg23944.html). The percentage scored for each fraction is computed as the number of hits assigned to each fraction times 100 divided by the total number of PIs taken (usually 1010, but this number can vary).

A height rule is applied, so that scoring the green/brown/bare fraction from point intercept hits of the uppermost vegetation/substrate stratum overrides the others. That is, a dead tree overrides a living shrub beneath and vice versa. Similarly, any vegetation cover overrides substrate coding, etc. This means for each of the intercepts, there is a single coding.

Arguments:

- `ground_fractional` : When set to 'TRUE' it computes fractional cover exclusively for Ground Cover. That is, fractional cover is computed only for grasses (hummock, tussock, other); sedge; rush; forb; fern; and vine plant growth forms. Presently, cryptogam cover is excluded, and included in the non-photosynthetic fraction instead. The default for this argument is 'FALSE'.
- `in_canopy_sky` : With the default value (`in_canopy_sky = FALSE`), only the substrate is considered for those hits. This argument applies only to regular fractional cover (i.e. for `ground_fractional = FALSE`), as trees are excluded in the green fraction for ground fractional cover by default.

Occasionally, substrate type was not collected (i.e. 'NC') or could not be assigned to one of the above categories (i.e. 'Unknwn'). In these cases, if there was no vegetation cover above those points percent cover is scored as an 'NA' fraction.

The `fractional_cover` function returns a data frame. In this data frame rows denote plots, columns denote fractions (i.e. bare, brown, green, and NA), and values are cover percentages.

In this section we will explore:

- The Latitudinal Pattern in Proportional Vegetation Cover (for a random subest of 200 sites).
- Temporal Variation in fractional cover: Explore, display, and assess (for 5 sites visited twice).

Latitudinal Pattern in Proportional Vegetation Cover

In this seection we will follow these steps:

- Call the `fractional_cover` function on the extracted point intercept data. This calculation may take a few minutes for all AusPlots, so for this example we will work with a random subset of 200 randomly drawn sites.
- Plot the Latitudinal Pattern in Proportional Vegetation Cover (here we use the 'Proportion of Bare Ground'). To do this, we first enrich the dataset with additional variables including: 'Plot' (identifier for each Site-Visit combination), 'bioregion.f', 'longitude', and 'latitude'.
- Fit a Quadratic Model to the data and examine its Fit, as there appears to be a humpbacked relationship in the previous plot (higher proportionof bare ground in the arid inland at mid-latitudes).

```
# First, we call the fractional_cover function on the extracted point intercept data.
# NOTE: Calculation may take a few minutes for all AusPlots, so for this example
# we will pull out a subset of 200 randomly drawn sites to work with.
# The sets site composition will differ each time the script is run,
# as they are random subsets.


# Compute Fractional Cover using function 'fractional_cover`
# --------------------------------------------------------
AP.200Locs.FC =
fractional_cover(AP.data$veg.PI[AP.data$veg.PI$site_location_name %in%
                                sample(AP.data$site.info$site_location_name, 200), ])
# To use the Full Data set substitute the command above by this one:
# fractional_cover(AP.data$veg.PI[AP.data$veg.PI$site_location_name %in%
#                           AP.data$site.info$site_location_name, ])
#AP.200Locs.FC = na.omit(AP.200Loc.FC)
head(AP.200Locs.FC)
```

```
##                          site_unique  bare brown green  NA.
## NSABBS0006-58557 NSABBS0006-58557   3.17 32.67 64.06  0.1
## NSABHC0001-53596 NSABHC0001-53596  31.26 42.24 26.51  0.0
## NSABHC0005-53600 NSABHC0005-53600  20.30 32.77 46.93  0.0
## NSABHC0006-53601 NSABHC0006-53601  22.87 26.53 50.59  0.0
## NSABHC0008-53603 NSABHC0008-53603  37.65 30.00 32.35  0.0
## NSABHC0010-53605 NSABHC0010-53605  75.27  9.89 14.84  0.0
```

```
# Enrich with: Bioregion, Latitude, and Longitude
# -----------------------------------------------
AP.200Locs.FC  = merge(AP.200Locs.FC, AP.data$site.info, by="site_unique")[,c("site_unique",
"bare", "brown", "green", "NA.", "bioregion.f", "longitude", "latitude")]
AP.200Locs.FC = na.omit(AP.200Locs.FC)
head(AP.200Locs.FC)
```

```
##          site_unique  bare brown green NA. bioregion.f longitude  latitude
## 1 NSABBS0006-58557  3.17 32.67 64.06 0.1         BBS  149.2673 -30.51027
## 2 NSABHC0001-53596 31.26 42.24 26.51 0.0         BHC  141.4338 -31.89335
## 3 NSABHC0005-53600 20.30 32.77 46.93 0.0         BHC  141.9405 -31.84496
## 4 NSABHC0006-53601 22.87 26.53 50.59 0.0         BHC  141.7823 -31.88421
## 5 NSABHC0008-53603 37.65 30.00 32.35 0.0         BHC  141.5288 -31.88592
## 6 NSABHC0010-53605 75.27  9.89 14.84 0.0         BHC  141.7003 -31.07672
```

```
summary(AP.200Locs.FC)
```

```
##  site_unique              bare              brown              green
## Length:225          Min.   : 0.10   Min.   : 1.19   Min.   : 0.10
## Class :character    1st Qu.:10.79   1st Qu.:20.89   1st Qu.:21.58
## Mode  :character    Median :22.48   Median :30.40   Median :37.72
##                     Mean   :29.21   Mean   :31.26   Mean   :39.45
##                     3rd Qu.:43.47   3rd Qu.:40.00   3rd Qu.:52.45
##                     Max.   :98.71   Max.   :70.20   Max.   :96.34
##
##       NA.             bioregion.f    longitude        latitude
## Min.   :0.00000   FLB    : 25   Min.   :114.6   Min.   :-37.41
## 1st Qu.:0.00000   SSD    : 21   1st Qu.:133.4   1st Qu.:-32.31
## Median :0.00000   MAC    : 17   Median :138.2   Median :-27.96
## Mean   :0.08044   MDD    : 17   Mean   :136.1   Mean   :-27.02
## 3rd Qu.:0.00000   BHC    : 13   3rd Qu.:141.0   3rd Qu.:-23.51
## Max.   :5.05000   COO    : 13   Max.   :150.4   Max.   :-12.22
##                   (Other):119
```

```
names(AP.200Locs.FC)
```

```
## [1] "site_unique" "bare"        "brown"        "green"        "NA."
## [6] "bioregion.f" "longitude"    "latitude"
```

```
# Plot out the continental relationship between Fractional Cover
# --------------------------------------------------------------
# Here we use the 'Proportion of Bare Ground' & Latitude

# Plot the relationship between Proportion of Bare Ground (with no kind of vegetation cover a
bove) and Latitude.
par(mfrow=c(1,1))
plot(bare ~ latitude, data=AP.200Locs.FC, pch=20, bty="l")
```
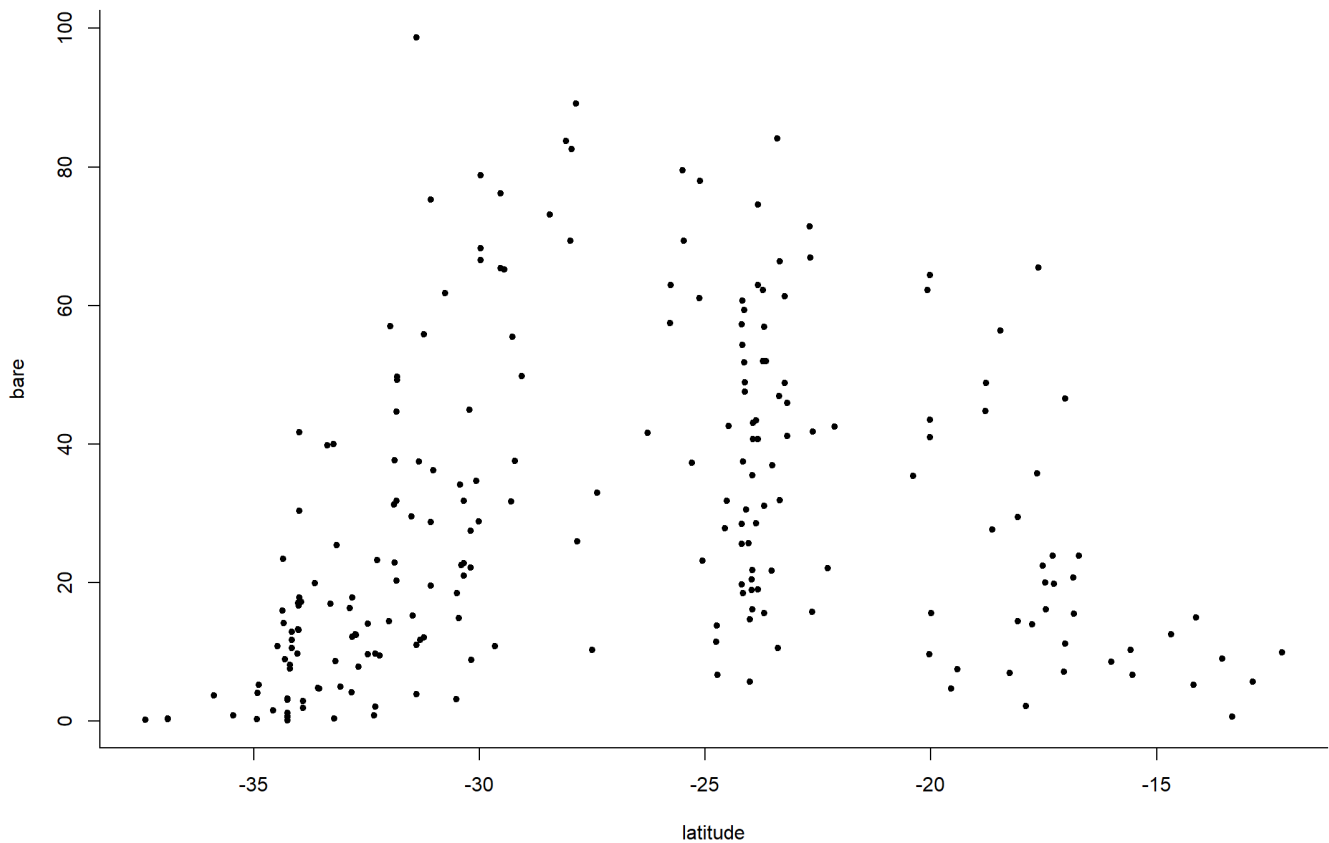
```
# Quadratic LM of Continental Relationship between Bare Ground Fractional Cover & Latitude
# --------------------------------------------------------------------------------

# Fit & Examine as Quadratic Linear Model the Continental Relationship between Bare Ground Fr
actional Cover & Latitude
AP.200Locs.FC.lm = lm(bare ~ latitude + I(latitude^2), data=AP.200Locs.FC)
summary(AP.200Locs.FC.lm)
```

```
##
## Call:
## lm(formula = bare ~ latitude + I(latitude^2), data = AP.200Locs.FC)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.669 -12.595  -1.936  10.693  70.072
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -166.42363   22.36427  -7.441 2.16e-12 ***
## latitude       -16.93015    1.79654  -9.424  < 2e-16 ***
## I(latitude^2)   -0.34133    0.03468  -9.843  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.8 on 222 degrees of freedom
## Multiple R-squared:  0.3241, Adjusted R-squared:  0.318
## F-statistic: 53.22 on 2 and 222 DF,  p-value: < 2.2e-16
```

```
# Predict values from Model Fit
pred.df = data.frame(latitude=seq(from=min(AP.200Locs.FC$latitude), to=max(AP.200Locs.FC$lati
tude), length.out=50))
pred.df$pred = predict(AP.200Locs.FC.lm, pred.df)

# Plot Predicted Values from Model Fit on Graph with Continental Relationship between Bare Gr
ound Fractional Cover & Latitude
plot(bare ~ latitude, data=AP.200Locs.FC, pch=20, bty="n")
points(pred.df$latitude, pred.df$pred, type="l", lwd=2, col="darkblue")
```
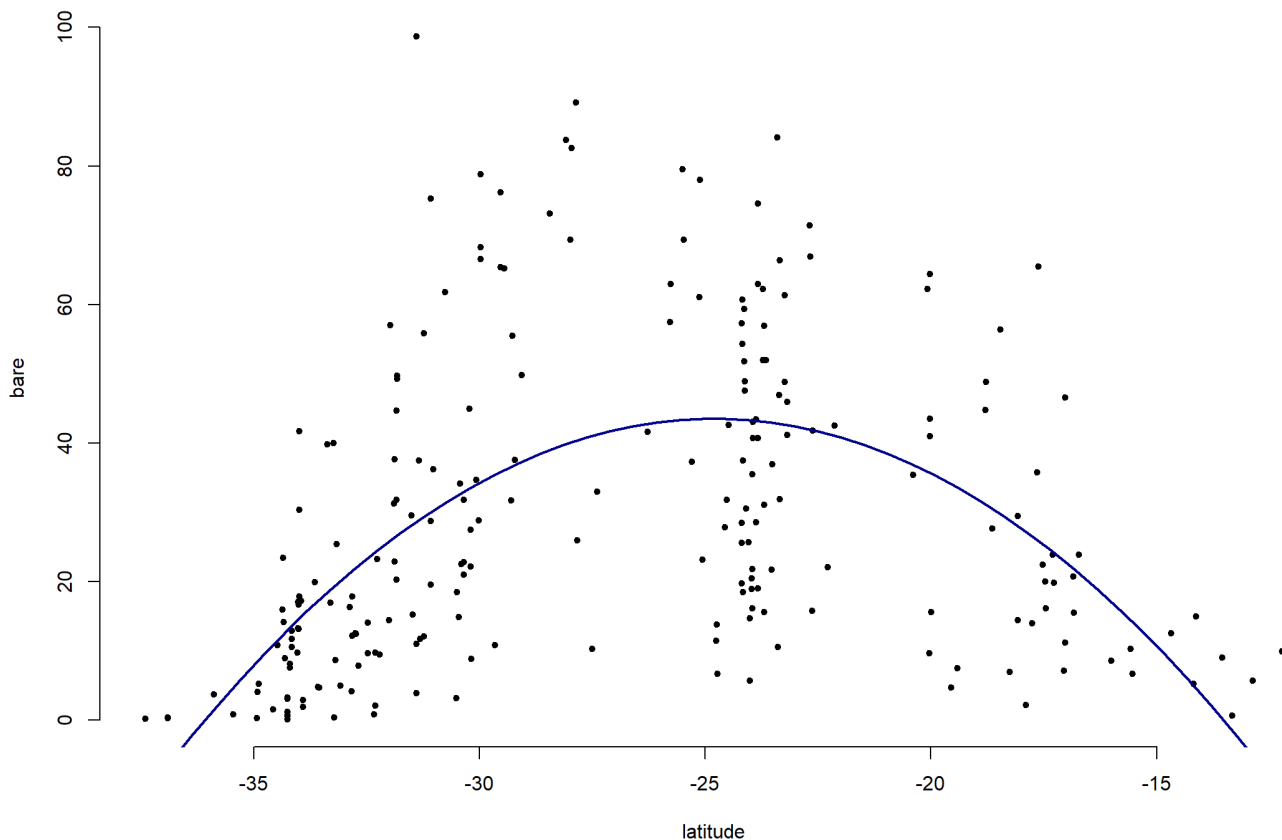

Temporal Variation in Fractional Cover: Explore, display, and assess (for 5 sites visited twice)

In the second section on Fractional Cover, we first Identify Sites that have been sampled more than once over time (in the 5 most sampled bioregions). Then we visually compare the Temporal Variation in Fractional Cover in these sites using Piecharts.

```
# Find Sites Sampled > 1 time
# --------------------------
# Extract Sites Names
AP.200Locs.FC.locs = sub("\\-.*", "", AP.200Locs.FC$site_unique)
length(AP.200Locs.FC.locs)
```

```
## [1] 225
```

```
# Calculate the Sample Frequency of each Site
AP.200Locs.FC.locs.cnt = count(AP.200Locs.FC.locs)
dim(AP.200Locs.FC.locs.cnt)
```

```
## [1] 195   2
```

```
# Find Sites with > 1 Samples (in veg.IP)
AP.200Locs.FC.Resampled.locs.cnt = AP.200Locs.FC.locs.cnt[AP.200Locs.FC.locs.cnt$freq > 1,]
dim(AP.200Locs.FC.Resampled.locs.cnt)
```

```
## [1] 30  2
```

```
AP.200Locs.FC.Resampled.locs.cnt = AP.200Locs.FC.Resampled.locs.cnt$x
length(AP.200Locs.FC.Resampled.locs.cnt)
```

```
## [1] 30
```

```
# Subset the Resampled Sites (i.e. with 'freq' > 1)
# --------------------------------------------------
# Extract AP.200Locs.FC subset for Sites with > 1 Samples (in veg.IP)
AP.200Locs.FC.Resampled.Locs = AP.200Locs.FC[(AP.200Locs.FC.locs %in% AP.200Locs.FC.Resample
d.locs.cnt),]
#AP.200Locs.FC.Resampled.Locs
dim(AP.200Locs.FC.Resampled.Locs) # Number of Samples taken in re-visited sites.
```

```
## [1] 60  8
```

```
#Number of Re-visited sites is half this, as each site was sampled twice (so far).

# Add Year (Started) Sampling of Site-Visit Pair
# ----------------------------------------------
# Need to specify 'AP.BioregTop5.l$site.info[,c("site_unique","visit_start_date")]' to avoid
 duplicate columns
AP.200Locs.FC.Resampled.Locs  = merge(AP.200Locs.FC.Resampled.Locs, AP.BioregTop5.l$site.info
[,c("site_unique","visit_start_date")],
                                                      by="site_unique")[,c(names(AP.200Loc
s.FC.Resampled.Locs),                                            "visit_start_date")]
names(AP.200Locs.FC.Resampled.Locs)
```

```
## [1] "site_unique"      "bare"             "brown"
## [4] "green"            "NA."              "bioregion.f"
## [7] "longitude"        "latitude"         "visit_start_date"
```

```
#AP.200Locs.FC.Resampled.Locs$visit_start_date
#substr(AP.200Locs.FC.Resampled.Locs$visit_start_date,1,4)
AP.200Locs.FC.Resampled.Locs$site_unique.Yr = paste( AP.200Locs.FC.Resampled.Locs$site_uniqu
e,
                                                      substr(AP.200Locs.FC.Resampled.Locs$visit_start
_date,1,4),
                                                      sep="." )
head(AP.200Locs.FC.Resampled.Locs$site_unique.Yr)
```

```
## [1] "QDASSD0002-53757.2014" "QDASSD0002-57622.2015" "QDASSD0005-56914.2014"
## [4] "QDASSD0005-57625.2015" "QDASSD0006-56915.2014" "QDASSD0006-57626.2015"
```
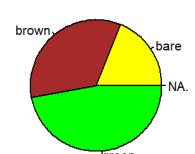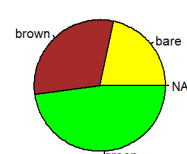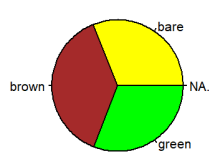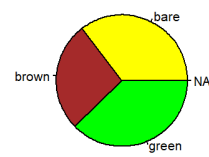
```
# Plot Pies for the first 5 Resampled Sites (i.e. 10 Site-Visit pairs)
# ---------------------------------------------------------------------
# Order dataframe to Plot Site-Visit pairs in the appropriate order
AP.200Locs.FC.Resampled.Locs = AP.200Locs.FC.Resampled.Locs[order(AP.200Locs.FC.Resampled.Loc
s$site_unique),]
head(AP.200Locs.FC.Resampled.Locs)
```

```
##           site_unique  bare brown green   NA. bioregion.f longitude  latitude
## 1 QDASSD0002-53757 62.97 16.93 20.10 0.00         SSD  138.3897 -23.82772
## 2 QDASSD0002-57622 40.69 29.50 29.80 0.00         SSD  138.3897 -23.82772
## 3 QDASSD0005-56914 56.93 24.36 16.83 1.88         SSD  138.4410 -23.69444
## 4 QDASSD0005-57625 31.02 38.06 30.91 0.00         SSD  138.4410 -23.69444
## 5 QDASSD0006-56915 43.37 38.22 18.42 0.00         SSD  138.0984 -23.87954
## 6 QDASSD0006-57626 28.51 23.86 47.62 0.00         SSD  138.0984 -23.87954
##        visit_start_date        site_unique.Yr
## 1 2014-05-01T00:00:00 QDASSD0002-53757.2014
## 2 2015-04-16T00:00:00 QDASSD0002-57622.2015
## 3 2014-05-02T00:00:00 QDASSD0005-56914.2014
## 4 2015-04-26T00:00:00 QDASSD0005-57625.2015
## 5 2014-05-03T00:00:00 QDASSD0006-56915.2014
## 6 2015-04-18T00:00:00 QDASSD0006-57626.2015
```

```
# Plot the Site-Visit pairs
par(mfcol=c(2,5))
for (site.visit.cnt in 1:10) {

    pie( x=as.numeric(AP.200Locs.FC.Resampled.Locs[site.visit.cnt,2:5]),
         col=c("yellow", "brown", "green", "white"),
         labels=names(AP.200Locs.FC.Resampled.Locs[2:5]),
         main=as.character(AP.200Locs.FC.Resampled.Locs[site.visit.cnt,"site_unique.Yr"]) )

} # for site.visit.cnt in 1:20 {
```

QDASSD0002-53757.2014 · QDASSD0005-56914.2014 · QDASSD0006-56915.2014 · QDASSD0008-56917.2014 · QDASSD0009-56918.2014

QDASSD0002-57622.2015 · QDASSD0005-57625.2015 · QDASSD0006-57626.2015 · QDASSD0008-57628.2015 · QDASSD0009-57629.2015
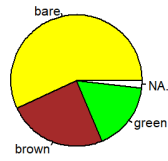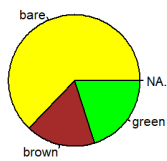
GROWTH FORM: `growth_form_table` function (for 5 most sampled bioregions)

The `growth_form_table` function in the `ausplotR` package can be used to generate occurrence matrices for plant growth forms in the AusPlots plots. The plant growth forms considered are those in the National Vegetation Information System (NVIS; Executive Steering Committee for Australian Vegetation Information, 2003).

The input for the `growth_form_table` function is a data frame of raw AusPlots point intercept data generated using the `get_ausplots function` . Three metrics can be selected to score species growth form:

- *Presence/Absence*: Set by the argument `m_kind = "PA"` .
- *Percent Cover*: Based on total frequency of hits. Set by the argument `m_kind = "percent_cover"` . This is the most useful and commonly used metric. It can be subsequently used in statistical analyses (e.g. MANOVA, Ordination, Classification, etc.) at continental scale where species turnover is too high for some methods to provide meaningful results.
- *Species Richness*: Set by the argument `m_kind = "richness"` . Note that when 'm_kind' is set to "richness" the rowSums of the occurrence matrix can be higher than the observed SR because in some occasions the same species is recorded with different growth forms in a plot and therefore the same species can count towards the weights for multiple growth forms.

When Percent Cover is used two types of cover type can be selected:

- *Projected Foliage Cover (PFC)*: Hits scored as 'in canopy sky' are removed. Set by the argument `cover_type = "PFC"` .
- *Opaque Canopy Cover (OCC)*: Hits scored as 'in canopy sky' are retained. Set by the argument `cover_type = "OCC"` .

The output of the `growth_form_table` function is a data frame. In this output data frame rows correspond to unique sites, columns to NVIS plant growth forms, and cell values are either Presence/Absences, Percent Cover, or Species Richness depending on the option chosen for the 'm_kind' argument.

REFERENCES
**NationalVegetation Information System, Version 6.0*
Executive Steering Committee for Australian Vegetation Information (ESCAVI)
Department of the Environment and Heritage, 2003
ISBN 0 642 54953 2.

In this section we will:

- Generate a Plant Growth Forms Percent Cover against Sites Matrix using the `growth_form_table` function.
- Enrich this Matrix with additional information (plot -site-visit-, bioregion, longitude, and latitude).
- Compute Summary Statics for each of the Growth Forms in the 5 most sampled Bioregions (slightly different to those produce by the `summary` function in the `base` package.
- Cluster (Hierachical Clustering) the Sites-Visits by Plant Growth Forms Percent Cover, colouring the resulting tree branches by bioregion.

Clustering Results:

- The first Site-Visit (NTAGFU0007-53654) is very different to the rest
- The dendrogram shows clusters formed by single Bioregions at low level; however, at higher-level clusters are composed by Sites-Visits from different Bioregions.

```
# Generate the Growth Form by Site-Visit Matrix
# =============================================
AP.BRTop5.GrowthFormBYSites = growth_form_table(AP.BioregTop5.l$veg.PI,
m_kind="percent_cover", cover_type="PFC")  # % Cover
dim(AP.BRTop5.GrowthFormBYSites) # No of rows and cols in Matrix: Sites x Growth Forms
```

```
## [1] 225  18
```

```
head(AP.BRTop5.GrowthFormBYSites)
```

```
##                    Bryophyte  Chenopod  Epiphyte Fern        Forb    Fungus
## NSAMDD0001-56965           0 19.537815 0.0000000    0 83.1932773 0.4201681
## NSAMDD0002-56952           0  3.024911 0.0000000    0  0.1779359 0.0000000
## NSAMDD0003-56968           0 24.635036 0.0000000    0  9.3065693 0.0000000
## NSAMDD0004-56953           0 45.194805 0.0000000    0 31.9480519 0.0000000
## NSAMDD0005-56969           0 12.923077 0.0000000    0 12.0000000 0.0000000
## NSAMDD0006-56954           0 79.945799 0.2710027    0  3.2520325 0.0000000
##                    Grass.tree Heath.shrub Hummock.grass        NC Rush Sedge
## NSAMDD0001-56965            0           0       0.00000 2.7310924    0     0
## NSAMDD0002-56952            0           0      31.13879 0.0000000    0     0
## NSAMDD0003-56968            0           0       0.00000 0.0000000    0     0
## NSAMDD0004-56953            0           0       0.00000 0.0000000    0     0
## NSAMDD0005-56969            0           0       0.00000 0.9230769    0     0
## NSAMDD0006-56954            0           0       0.00000 0.0000000    0     0
##                        Shrub Shrub.Mallee Tree.Mallee Tree.Palm
## NSAMDD0001-56965   0.0000000    0.0000000   0.0000000  0.000000
## NSAMDD0002-56952   1.6014235    0.5338078  76.8683274  0.000000
## NSAMDD0003-56968  82.8467153    0.0000000   0.3649635  1.094891
## NSAMDD0004-56953   5.1948052    0.0000000   0.0000000 28.831169
## NSAMDD0005-56969   4.0000000    0.0000000  76.6153846  0.000000
## NSAMDD0006-56954   0.8130081    0.0000000   0.0000000 32.249322
##                    Tussock.grass Vine
## NSAMDD0001-56965       1.8907563    0
## NSAMDD0002-56952       0.3558719    0
## NSAMDD0003-56968       0.0000000    0
## NSAMDD0004-56953       1.0389610    0
## NSAMDD0005-56969       0.3076923    0
## NSAMDD0006-56954       0.5420054    0
```

```
# Enrich DF
# =========

# Create a 'site_unique' variable in Growth Form by Site-VisitTable to relate both datasets
# ------------------------------------------------------------------------------------------
AP.BRTop5.GrowthFormBYSites$site_unique = rownames(AP.BRTop5.GrowthFormBYSites)

# Add: Bioregion, Longitude, Latitude
# ----------------------------------
# Both DF have different number of rows (again!)
dim(AP.BRTop5.GrowthFormBYSites)
```

```
## [1] 225  19
```

```
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231  44
```

```
# Enrich with: Bioregion, Latitude, and Longitude
AP.BRTop5.GrowthFormBYSites = merge(AP.BRTop5.GrowthFormBYSites, AP.BioregTop5.l$site.info,
 by="site_unique")[,c(names(AP.BRTop5.GrowthFormBYSites),

"bioregion.f", "longitude", "latitude")]
AP.BRTop5.GrowthFormBYSites = na.omit(AP.BRTop5.GrowthFormBYSites)
#head(AP.BRTop5.GrowthFormBYSites)
summary(AP.BRTop5.GrowthFormBYSites)
```

```
##    Bryophyte          Chenopod          Epiphyte           Fern
##  Min.   :0.00000   Min.   :  0.00   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.:0.00000   1st Qu.:  0.00   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median :0.00000   Median :  1.60   Median : 0.0000   Median : 0.0000
##  Mean   :0.02563   Mean   : 14.42   Mean   : 0.1215   Mean   : 0.3888
##  3rd Qu.:0.00000   3rd Qu.: 22.01   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :5.50459   Max.   :100.00   Max.   :14.2857   Max.   :20.6612
##      Forb             Fungus            Grass.tree         Heath.shrub
##  Min.   : 0.000   Min.   :0.000000   Min.   : 0.0000   Min.   :0.00000
##  1st Qu.: 2.874   1st Qu.:0.000000   1st Qu.: 0.0000   1st Qu.:0.00000
##  Median : 8.108   Median :0.000000   Median : 0.0000   Median :0.00000
##  Mean   :15.421   Mean   :0.001867   Mean   : 0.6295   Mean   :0.06194
##  3rd Qu.:20.614   3rd Qu.:0.000000   3rd Qu.: 0.0000   3rd Qu.:0.00000
##  Max.   :94.268   Max.   :0.420168   Max.   :50.1805   Max.   :7.20412
##  Hummock.grass          NC               Rush              Sedge
##  Min.   :  0.00   Min.   :0.00000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.:  0.00   1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median :  0.00   Median :0.00000   Median : 0.0000   Median : 0.0000
##  Mean   : 12.32   Mean   :0.06185   Mean   : 0.1589   Mean   : 0.6737
##  3rd Qu.:  9.47   3rd Qu.:0.00000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :183.86   Max.   :2.80374   Max.   :12.3487   Max.   :29.2653
##      Shrub           Shrub.Mallee       Tree.Mallee        Tree.Palm
##  Min.   :  0.000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.:  2.899   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 13.126   Median : 0.0000   Median : 0.0000   Median : 0.7634
##  Mean   : 22.285   Mean   : 0.8482   Mean   : 7.6667   Mean   :15.9733
##  3rd Qu.: 34.064   3rd Qu.: 0.0000   3rd Qu.: 0.2503   3rd Qu.:26.3375
##  Max.   :100.168   Max.   :37.7863   Max.   :91.9786   Max.   :99.1747
##  Tussock.grass        Vine          site_unique        bioregion.f
##  Min.   :  0.00   Min.   :0.0000   Length:225         FLB:48
##  1st Qu.:  1.22   1st Qu.:0.0000   Class :character   GFU:41
##  Median : 12.08   Median :0.0000   Mode  :character   MDD:50
##  Mean   : 22.90   Mean   :0.2121                      SSD:46
##  3rd Qu.: 40.17   3rd Qu.:0.0000                      STP:40
##  Max.   : 99.05   Max.   :7.7670
##    longitude         latitude
##  Min.   :132.9   Min.   :-35.08
##  1st Qu.:136.3   1st Qu.:-33.09
##  Median :138.2   Median :-28.43
##  Mean   :138.3   Mean   :-27.35
##  3rd Qu.:139.3   3rd Qu.:-23.83
##  Max.   :145.8   Max.   :-14.05
```

```
names(AP.BRTop5.GrowthFormBYSites)
```

```
##  [1] "Bryophyte"      "Chenopod"       "Epiphyte"       "Fern"
##  [5] "Forb"           "Fungus"         "Grass.tree"     "Heath.shrub"
##  [9] "Hummock.grass"  "NC"             "Rush"           "Sedge"
## [13] "Shrub"          "Shrub.Mallee"   "Tree.Mallee"    "Tree.Palm"
## [17] "Tussock.grass"  "Vine"           "site_unique"    "bioregion.f"
## [21] "longitude"      "latitude"
```

```
# Summary Statistics for Each Growth Form
# =====================================
AP.BRTop5.GFBYSites.DescStats = data.frame(
  Min = apply(AP.BRTop5.GrowthFormBYSites[,1:(dim(AP.BRTop5.GrowthFormBYSites)[2]-4)], 2,
min), # Minimum
  Med = apply(AP.BRTop5.GrowthFormBYSites[,1:(dim(AP.BRTop5.GrowthFormBYSites)[2]-4)], 2,
median), # Median
  Max = apply(AP.BRTop5.GrowthFormBYSites[,1:(dim(AP.BRTop5.GrowthFormBYSites)[2]-4)], 2,
max), # Maximum
  Mean = apply(AP.BRTop5.GrowthFormBYSites[,1:(dim(AP.BRTop5.GrowthFormBYSites)[2]-4)], 2,
mean), # Mean
  SD = apply(AP.BRTop5.GrowthFormBYSites[,1:(dim(AP.BRTop5.GrowthFormBYSites)[2]-4)], 2,
sd) # Standard Deviation
  )
AP.BRTop5.GFBYSites.DescStats = round(AP.BRTop5.GFBYSites.DescStats, 2)
AP.BRTop5.GFBYSites.DescStats
```

```
##                 Min   Med     Max   Mean     SD
## Bryophyte         0  0.00    5.50   0.03   0.37
## Chenopod          0  1.60  100.00  14.42  22.22
## Epiphyte          0  0.00   14.29   0.12   0.99
## Fern              0  0.00   20.66   0.39   1.88
## Forb              0  8.11   94.27  15.42  18.91
## Fungus            0  0.00    0.42   0.00   0.03
## Grass.tree        0  0.00   50.18   0.63   4.60
## Heath.shrub       0  0.00    7.20   0.06   0.58
## Hummock.grass     0  0.00  183.86  12.32  25.94
## NC                0  0.00    2.80   0.06   0.32
## Rush              0  0.00   12.35   0.16   1.18
## Sedge             0  0.00   29.27   0.67   2.86
## Shrub             0 13.13  100.17  22.29  24.57
## Shrub.Mallee      0  0.00   37.79   0.85   4.40
## Tree.Mallee       0  0.00   91.98   7.67  18.06
## Tree.Palm         0  0.76   99.17  15.97  25.47
## Tussock.grass     0 12.08   99.05  22.90  26.19
## Vine              0  0.00    7.77   0.21   0.78
```

```
# Create and Plot a Dendogram of the Sites-Visits Clustered by Growth Forms
# =========================================================================

# Add rownames to be used as Leaves Names
rownames(AP.BRTop5.GrowthFormBYSites) = AP.BRTop5.GrowthFormBYSites$site_unique

# Create Dendogram
AP.BRTop5.GFBYSites.dend = as.dendrogram(hclust(dist(AP.BRTop5.GrowthFormBYSites[,1: (dim(AP.
BRTop5.GrowthFormBYSites)[2]-4)]), "average"))

# Color the Leaves by Bioregion
# NOTE: The most sampled bioregions might change as new data is added. If so, bioregions code
s below should be revised.
# The (currently) most common codes correspond to: MDD (Murry Darling Depression),
# SSD (Simpson Strzelecki Dunefields), GFU (Gulf Fall and Uplands), STP (Stony
# Plains), PIL (Pilbara), FLB (Flinders Lofty Block).
# For example, here PIL is commented out 'cos when writting the current version of
# this tutorial it wasn't one of the 5 most sampled bioregions. However, in the
# previous version of this tutorial PIL was included, and FLB wasn't among the 5
# most sampled bioregions.

AP.BRTop5.GrowthFormBYSites$bioregion.col.f = AP.BRTop5.GrowthFormBYSites$bioregion.f
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioreg
ion.col.f) == "GFU"] = "darkgreen"
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioreg
ion.col.f) == "MDD"] = "magenta"
#levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBY#Sites$bior
egion.col.f) == "PIL"] = "red"
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioreg
ion.col.f) == "FLB"] = "red"
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioreg
ion.col.f) == "SSD"] = "yellow"
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioreg
ion.col.f) == "STP"] = "cyan"
dend.colors = as.character(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)
#dend.colors
dend.colors = dend.colors[order.dendrogram(AP.BRTop5.GFBYSites.dend)]
#dend.colors
labels_colors(AP.BRTop5.GFBYSites.dend) = dend.colors
# Plot Dendrogram
par(mfrow=c(1,1))
plot(AP.BRTop5.GFBYSites.dend,
     main="Dendogram of the Sites-Visits Clustered by Growth Forms, with leaves coloured by B
ioregion")
```
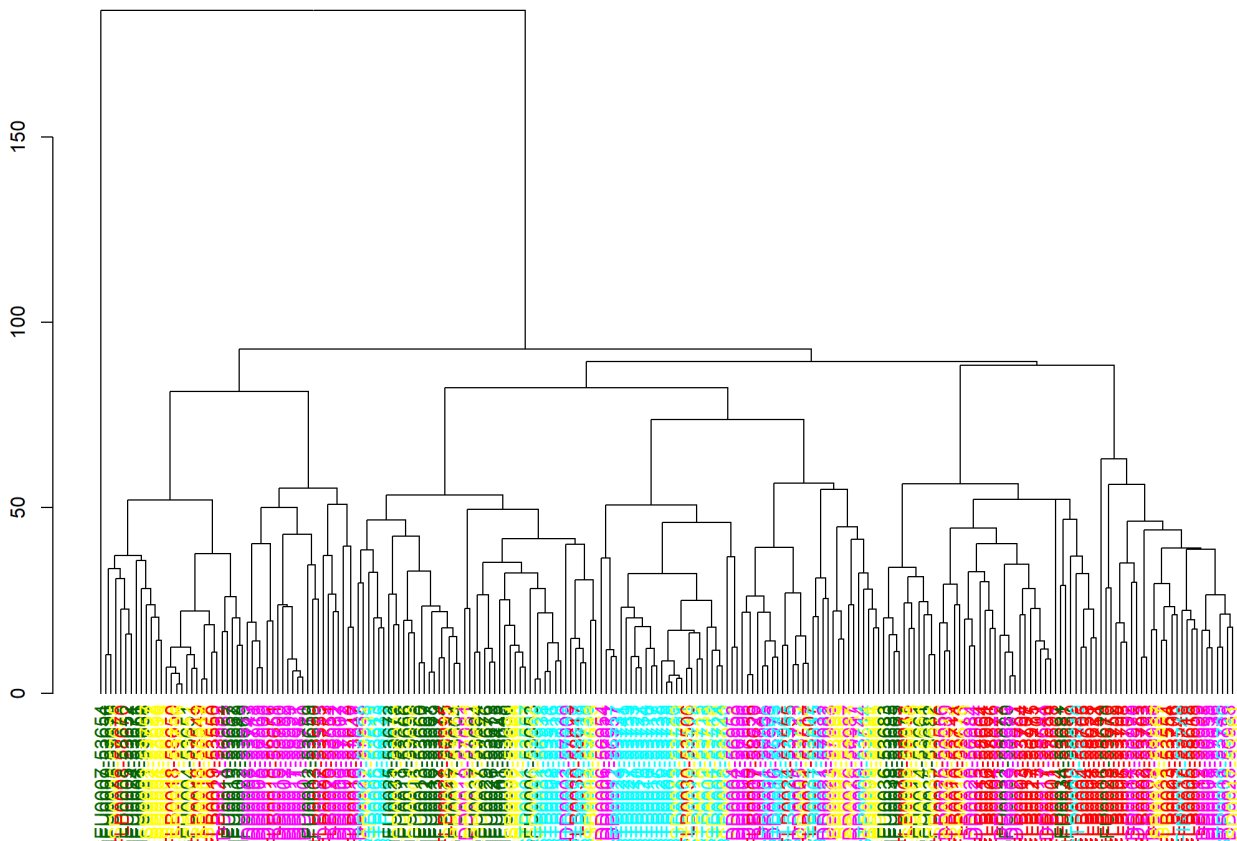
Dendogram of the Sites-Visits Clustered by Growth Forms, with leaves coloured by Bioregion

VEGETATION COVER BY GROWTH FORM AND/OR HEIGHT: `single_cover_value` function (for 5 most sampled bioregions).

The `single_cover_value` function in the `auplotsR` package calculates Vegetation Cover Values for particular Growth Form Types and/or Height Thresholds per Site from Raw AusPlots Vegetation Point Intercept data. The `growth_form_table` function can also be used to calculate Cover Values for all Vegetation Growth Form Types; however, `single_cover_value` can perform these computations for:

- Particular vegetation growth form types (i.e. for individual growth forms or any combination of growth form types).
- Vegetation higher that a specified height threshold
- Vegetation with any combination of growth form types and minimum height

Specifically `single_cover_value` takes the following inputs via its arguments:

- `veg.PI` : Raw Vegetation Point Intercept data from AusPlots. A veg.PI data frame generated by the `get_ausplots` function (see above).
- `in_canopy_sky` : Method used to calculate Cover. A logical value that indicates whether to use in 'canopy sky hits' (i.e. calculate 'opaque canopy cover') or 'projected foliage cover'. The default value, 'FALSE', calculates 'projected foliage cover'. To calculate 'opaque canopy cover' the argument must be set to 'TRUE'.
- `by.growth_form` : Whether to calculate Cover for a Subset by Growth Form type. A logical value that indicates whether to subset by growth form type. The default, 'TRUE', calculates cover for the growth form types specified in the argument 'my.growth_forms' (see next). If set to 'FALSE', cover calculations are conducted only for the vegetation sub-set by a provided Minimum Height Threshold.
- `my.growth_forms` : Growth Form Types used to Subset Data used for the Cover Calculations. A character vector specifying the growth form types to subset the data used for the cover calculations. Any combination of growth form types can be used. The default, 'c("Tree/Palm", "Tree Mallee")', is set to represent trees. It applies only when 'by.growth_form=TRUE'; otherwise, this argument is ignored and only height sub-setting is applied.

- `min.height` : Minimum Height Threshold used to Subset Data used for the Cover Calculations. A numeric value indicating the minimum height (in metres) of the vegetation to be included in the subset of the data used for the cover calculations. A height must be always provided. The default, '5', is set up for a cover of trees. It can be set to '0' to ignore height and thus include any plant hit. If set to a 'negative number', it will return nonsensical output.

The `single_cover_value` function returns a data frame with two columns. The data frame rows correspond to unique sites, while the two columns correspond to the unique site and the percentage cover for the requested subset of vegetation (e.g. "Tree/Palm" higher than '5' metres).

When `by.growth_form = FALSE` and `min.height = 0` , the output is nearly the same as the green cover fraction returned by the `fractional_cover` function (see above). The values can differ because 'fractional_cover' applies a 'height rule' in which the highest intercept at a given point is taken, whereas 'single_cover_value' finds any green cover. For example, when dead trees overhang green understorey the values returned by both functions can differ. For general cover purposes, using 'fractional_cover' is recommended. 'single_cover_value' is best suited to calculate cover subset by height and growth form. Next, several examples of how to compute, manipulate, and visualise 'Single' Vegetation Cover Fraction (VCF) data are presented. The examples cover different scenarios for subsetting the input vegetation point intercept data frame prior to the calculation of the corresponding VCF. These include:

- Subsetting only by Height
- Subsetting only by Taxonomoy
- Subsetting by Height and Taxonomy

Subsetting by Height only

```
# Subsetting by Height only
# ************************

# Compute Single Cover Values Tables
# ===================================
# Any green vegeation (i.e. >= 0m in height)
veg.cover.gt0 = single_cover_value(AP.BioregTop5.l$veg.PI, by.growth_form=FALSE, min.height=0
)
# Any green vegeation >= 2m in height
veg.cover.gt2 = single_cover_value(AP.BioregTop5.l$veg.PI, by.growth_form=FALSE, min.height=2
)


# Combine all Tables into a Sigle Data Frame
# ==========================================
# Create a data frame containing all the Vegetation Cover Fractions
AP.BioregTop5.VCF.df = data.frame(site_unique=veg.cover.gt0$site_unique, VCF.gt0=veg.cover.gt
0$percentCover, VCF.gt2=veg.cover.gt2$percentCover)
head(AP.BioregTop5.VCF.df)
```

```
##          site_unique VCF.gt0 VCF.gt2
## 1 NSAMDD0001-56965    45.74    0.00
## 2 NSAMDD0002-56952    55.45   39.21
## 3 NSAMDD0003-56968    47.52    4.65
## 4 NSAMDD0004-56953    35.05   10.99
## 5 NSAMDD0005-56969    31.29   23.76
## 6 NSAMDD0006-56954    34.95   11.29
```

```
summary(AP.BioregTop5.VCF.df)
```

```
##           site_unique     VCF.gt0          VCF.gt2
##  NSAMDD0001-56965:  1   Min.   : 0.20    Min.   : 0.00
##  NSAMDD0002-56952:  1   1st Qu.:20.20    1st Qu.: 0.30
##  NSAMDD0003-56968:  1   Median :33.40    Median : 8.12
##  NSAMDD0004-56953:  1   Mean   :36.31    Mean   :12.97
##  NSAMDD0005-56969:  1   3rd Qu.:50.40    3rd Qu.:20.89
##  NSAMDD0006-56954:  1   Max.   :84.06    Max.   :70.89
##  (Other)         :219
```

```
# Enrich DF
# =========

# Compute Vegetation Cover Fractions for Height Ranges
# ------------------------------------------------------
AP.BioregTop5.VCF.df$VCF.0to2 = AP.BioregTop5.VCF.df$VCF.gt0 - AP.BioregTop5.VCF.df$VCF.gt2
head(AP.BioregTop5.VCF.df)
```

```
##        site_unique VCF.gt0 VCF.gt2 VCF.0to2
## 1 NSAMDD0001-56965   45.74    0.00    45.74
## 2 NSAMDD0002-56952   55.45   39.21    16.24
## 3 NSAMDD0003-56968   47.52    4.65    42.87
## 4 NSAMDD0004-56953   35.05   10.99    24.06
## 5 NSAMDD0005-56969   31.29   23.76     7.53
## 6 NSAMDD0006-56954   34.95   11.29    23.66
```

```
summary(AP.BioregTop5.VCF.df)
```

```
##           site_unique     VCF.gt0          VCF.gt2          VCF.0to2
##  NSAMDD0001-56965:  1   Min.   : 0.20    Min.   : 0.00    Min.   : 0.20
##  NSAMDD0002-56952:  1   1st Qu.:20.20    1st Qu.: 0.30    1st Qu.:12.28
##  NSAMDD0003-56968:  1   Median :33.40    Median : 8.12    Median :20.49
##  NSAMDD0004-56953:  1   Mean   :36.31    Mean   :12.97    Mean   :23.34
##  NSAMDD0005-56969:  1   3rd Qu.:50.40    3rd Qu.:20.89    3rd Qu.:32.67
##  NSAMDD0006-56954:  1   Max.   :84.06    Max.   :70.89    Max.   :68.02
##  (Other)         :219
```

```
# Add: Bioregion, Longitude, Latitude
# ------------------------------------

# Both DF have different number of rows
dim(AP.BioregTop5.VCF.df)
```

```
## [1] 225   4
```

```
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231  44
```

```
# Enrich with: Bioregion, Latitude, and Longitude
AP.BioregTop5.VCF.df  = merge(AP.BioregTop5.VCF.df, AP.BioregTop5.l$site.info, by="site_uniqu
e")[,c(names(AP.BioregTop5.VCF.df), "bioregion.f", "longitude", "latitude")]
AP.BioregTop5.VCF.df = na.omit(AP.BioregTop5.VCF.df)
head(AP.BioregTop5.VCF.df)
```

```
##        site_unique VCF.gt0 VCF.gt2 VCF.0to2 bioregion.f longitude
## 1 NSAMDD0001-56965   45.74    0.00    45.74         MDD  142.5602
## 2 NSAMDD0002-56952   55.45   39.21    16.24         MDD  142.6026
## 3 NSAMDD0003-56968   47.52    4.65    42.87         MDD  142.6041
## 4 NSAMDD0004-56953   35.05   10.99    24.06         MDD  142.5594
## 5 NSAMDD0005-56969   31.29   23.76     7.53         MDD  143.2039
## 6 NSAMDD0006-56954   34.95   11.29    23.66         MDD  143.1665
##    latitude
## 1 -34.18392
## 2 -34.20482
## 3 -34.20754
## 4 -34.16537
## 5 -33.65619
## 6 -33.66432
```

```
summary(AP.BioregTop5.VCF.df)
```

```
##           site_unique     VCF.gt0         VCF.gt2          VCF.0to2
## NSAMDD0001-56965:  1   Min.   : 0.20   Min.   : 0.00   Min.   : 0.20
## NSAMDD0002-56952:  1   1st Qu.:20.20   1st Qu.: 0.30   1st Qu.:12.28
## NSAMDD0003-56968:  1   Median :33.40   Median : 8.12   Median :20.49
## NSAMDD0004-56953:  1   Mean   :36.31   Mean   :12.97   Mean   :23.34
## NSAMDD0005-56969:  1   3rd Qu.:50.40   3rd Qu.:20.89   3rd Qu.:32.67
## NSAMDD0006-56954:  1   Max.   :84.06   Max.   :70.89   Max.   :68.02
## (Other)         :219
## bioregion.f   longitude        latitude
## FLB:48      Min.   :132.9   Min.   :-35.08
## GFU:41      1st Qu.:136.3   1st Qu.:-33.09
## MDD:50      Median :138.2   Median :-28.43
## SSD:46      Mean   :138.3   Mean   :-27.35
## STP:40      3rd Qu.:139.3   3rd Qu.:-23.83
##             Max.   :145.8   Max.   :-14.05
##
```

```
names(AP.BioregTop5.VCF.df)
```

```
## [1] "site_unique" "VCF.gt0"     "VCF.gt2"     "VCF.0to2"    "bioregion.f"
## [6] "longitude"   "latitude"
```

```r
# Graphical Visualisation
# =======================

# VCF 0 to 2m: Map with circle size = the Vegetation Cover Fraction for any green veg.
# -----------------------------------------------------------------------------------
# (i.e. >= 0m in height)
AP.BioregTop5.VCF.df.p1 =
ggplot(data=AP.BioregTop5.VCF.df, aes(x=longitude, y=latitude, colour=bioregion.f, fill=biore
gion.f), alpha =0.5) +
geom_point(aes(size=VCF.0to2), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# VCF 0 to 2m: Boxplot
# --------------------
AP.BioregTop5.VCF.df.p2 =
ggplot(AP.BioregTop5.VCF.df, aes(x=bioregion.f, y=VCF.0to2, color=bioregion.f)) +
geom_boxplot() +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(name="Bioregions (Top 5)", values = c("darkgreen", "magenta", "red", "yello
w", "cyan")) +
labs(x="Bioregion (5 most sampled)", y = "Vegetation Cover Fraction (0 to 2m)") +
theme(plot.title = element_text(hjust = 0.5))

# VCF >= 2m: Map with circle size = the Vegetation Cover Fraction for any green veg.
# ---------------------------------------------------------------------------------
# (i.e. >= 0m in height)
AP.BioregTop5.VCF.df.p3 =
ggplot(data=AP.BioregTop5.VCF.df, aes(x=longitude, y=latitude, colour=bioregion.f, fill=biore
gion.f), alpha =0.5) +
geom_point(aes(size=VCF.gt2), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# VCF  >= 2m: Boxplot
# -------------------
AP.BioregTop5.VCF.df.p4 =
ggplot(AP.BioregTop5.VCF.df, aes(x=bioregion.f, y=VCF.gt2, color=bioregion.f)) +
geom_boxplot() +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(name="Bioregions (Top 5)", values = c("darkgreen", "magenta", "red", "yello
w", "cyan")) +
labs(x="Bioregion (5 most sampled)", y = "Vegetation Cover Fraction (>= 2m)") +
theme(plot.title = element_text(hjust = 0.5))

# Plot both graphs
# ----------------
grid.arrange(AP.BioregTop5.VCF.df.p1, AP.BioregTop5.VCF.df.p2,
             AP.BioregTop5.VCF.df.p3, AP.BioregTop5.VCF.df.p4, nrow=2)
```
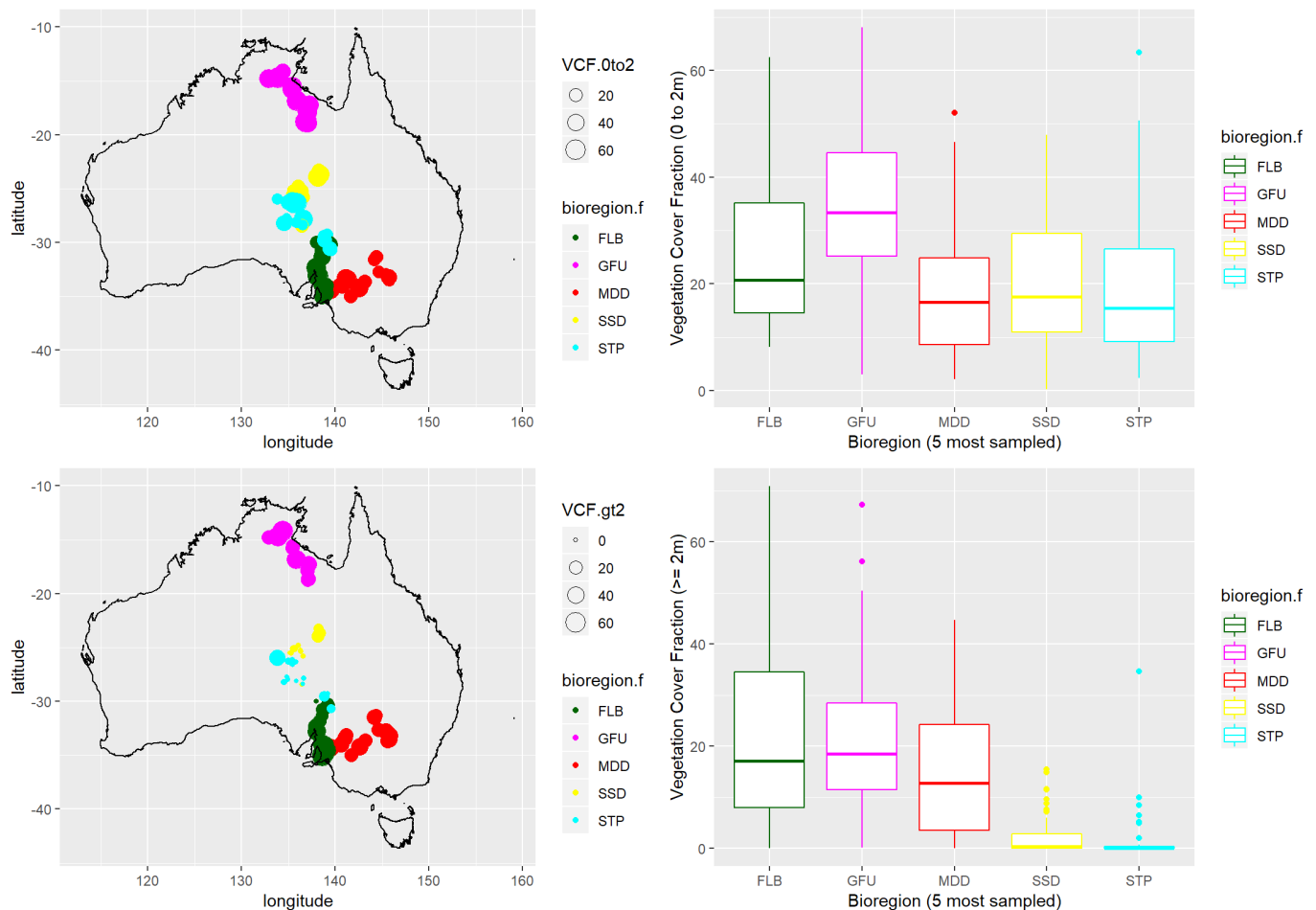
```
#grid.arrange(AP.BioregTop5.VCF.df.p1, AP.BioregTop5.VCF.df.p2,
#             AP.BioregTop5.VCF.df.p3, AP.BioregTop5.VCF.df.p4, ncol=2)
```

Subsetting by Taxonomy only

```
# Subsetting y Taxonomy only
# **************************

# Compute Single Cover Values Tables
# ===================================
# Trees (by default my.growth_forms=c("Tree/Palm", "Tree Mallee"))
veg.cover.trees = single_cover_value(AP.BioregTop5.l$veg.PI, min.height=0)
# Grasses (my.growth_forms=c("Hummock.grass", "Tussock.grass"))
veg.cover.grass = single_cover_value(AP.BioregTop5.l$veg.PI, by.growth_form=TRUE, my.growth_f
orms=c("Hummock grass", "Tussock grass"), min.height=0)


# Combine all Tables into a Sigle Data Frame
# ==========================================
# Create a data frame containing all the Vegetation Cover Fractions
AP.BioregTop5.VCF.df = data.frame(site_unique=veg.cover.trees$site_unique, VCF.trees=veg.cove
r.trees$percentCover, VCF.grass=veg.cover.grass$percentCover)
head(AP.BioregTop5.VCF.df)
```

```
##         site_unique VCF.trees VCF.grass
## 1 NSAMDD0001-56965      0.00      0.89
## 2 NSAMDD0002-56952     42.57     17.33
## 3 NSAMDD0003-56968      0.20      0.00
## 4 NSAMDD0004-56953     10.40      0.00
## 5 NSAMDD0005-56969     24.65      0.10
## 6 NSAMDD0006-56954     11.68      0.00
```

```
summary(AP.BioregTop5.VCF.df)
```

```
##          site_unique    VCF.trees       VCF.grass
##  NSAMDD0001-56965:  1   Min.   : 0.00   Min.   : 0.00
##  NSAMDD0002-56952:  1   1st Qu.: 0.00   1st Qu.: 0.69
##  NSAMDD0003-56968:  1   Median : 3.07   Median : 6.14
##  NSAMDD0004-56953:  1   Mean   :10.54   Mean   :13.10
##  NSAMDD0005-56969:  1   3rd Qu.:16.73   3rd Qu.:21.09
##  NSAMDD0006-56954:  1   Max.   :70.89   Max.   :67.43
##  (Other)         :219
```

```
# Enrich DF
# =========

# Add: Bioregion, Longitude, Latitude
# ----------------------------------

# Both DF have different number of rows
dim(AP.BioregTop5.VCF.df)
```

```
## [1] 225    3
```

```
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231   44
```

```
# Enrich with: Bioregion, Latitude, and Longitude
AP.BioregTop5.VCF.df  = merge(AP.BioregTop5.VCF.df, AP.BioregTop5.l$site.info, by="site_uniqu
e")[,c(names(AP.BioregTop5.VCF.df), "bioregion.f", "longitude", "latitude")]
AP.BioregTop5.VCF.df = na.omit(AP.BioregTop5.VCF.df)
head(AP.BioregTop5.VCF.df)
```

```
##         site_unique VCF.trees VCF.grass bioregion.f longitude  latitude
## 1 NSAMDD0001-56965      0.00      0.89         MDD  142.5602 -34.18392
## 2 NSAMDD0002-56952     42.57     17.33         MDD  142.6026 -34.20482
## 3 NSAMDD0003-56968      0.20      0.00         MDD  142.6041 -34.20754
## 4 NSAMDD0004-56953     10.40      0.00         MDD  142.5594 -34.16537
## 5 NSAMDD0005-56969     24.65      0.10         MDD  143.2039 -33.65619
## 6 NSAMDD0006-56954     11.68      0.00         MDD  143.1665 -33.66432
```

```
summary(AP.BioregTop5.VCF.df)
```

```
##          site_unique    VCF.trees       VCF.grass      bioregion.f
##  NSAMDD0001-56965:  1   Min.   : 0.00   Min.   : 0.00   FLB:48
##  NSAMDD0002-56952:  1   1st Qu.: 0.00   1st Qu.: 0.69   GFU:41
##  NSAMDD0003-56968:  1   Median : 3.07   Median : 6.14   MDD:50
##  NSAMDD0004-56953:  1   Mean   :10.54   Mean   :13.10   SSD:46
##  NSAMDD0005-56969:  1   3rd Qu.:16.73   3rd Qu.:21.09   STP:40
##  NSAMDD0006-56954:  1   Max.   :70.89   Max.   :67.43
##  (Other)         :219
##    longitude        latitude
##  Min.   :132.9   Min.   :-35.08
##  1st Qu.:136.3   1st Qu.:-33.09
##  Median :138.2   Median :-28.43
##  Mean   :138.3   Mean   :-27.35
##  3rd Qu.:139.3   3rd Qu.:-23.83
##  Max.   :145.8   Max.   :-14.05
##
```

```
names(AP.BioregTop5.VCF.df)
```

```
## [1] "site_unique" "VCF.trees"   "VCF.grass"   "bioregion.f" "longitude"
## [6] "latitude"
```

```
# Graphical Visualisation
# ========================

# Trees: Boxplot
# --------------
AP.BioregTop5.VCF.trees =
ggplot(AP.BioregTop5.VCF.df, aes(x=bioregion.f, y=VCF.trees, color=bioregion.f)) +
geom_boxplot() +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(name="Bioregions (Top 5)", values = c("darkgreen", "magenta", "red", "yello
w", "cyan")) +
labs(x="Bioregion (5 most sampled)", y = "Trees Cover Fraction (>= 0m)") +
theme(plot.title = element_text(hjust = 0.5))

# Grass: Boxplot
# --------------
AP.BioregTop5.VCF.grass =
ggplot(AP.BioregTop5.VCF.df, aes(x=bioregion.f, y=VCF.grass, color=bioregion.f)) +
geom_boxplot() +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(name="Bioregions (Top 5)", values = c("darkgreen", "magenta", "red", "yello
w", "cyan")) +
labs(x="Bioregion (5 most sampled)", y = "Grass Cover Fraction (>= 0m)") +
theme(plot.title = element_text(hjust = 0.5))

# Plot both graphs
# ----------------
grid.arrange(AP.BioregTop5.VCF.trees, AP.BioregTop5.VCF.grass, nrow=2)
```
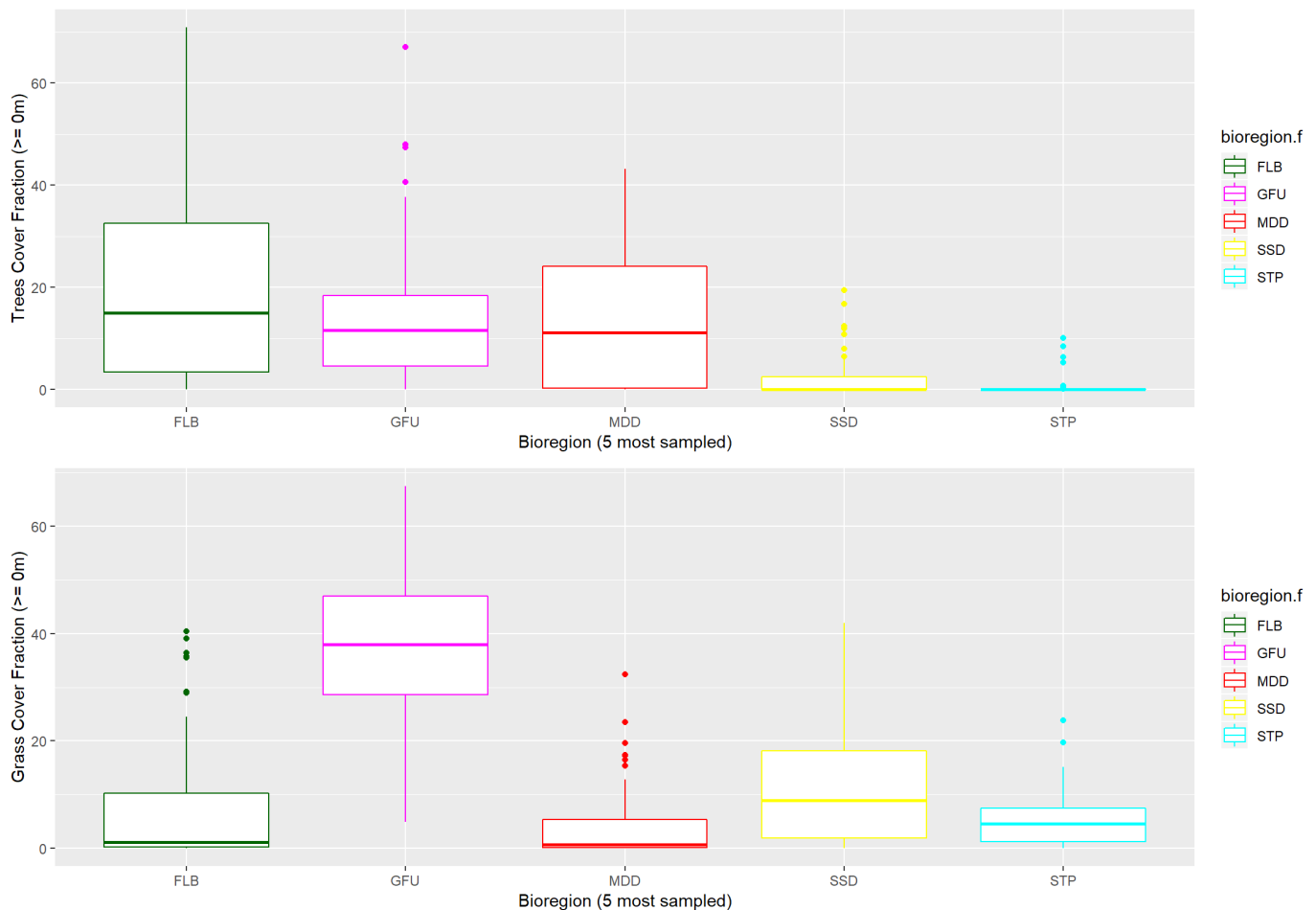
```
#grid.arrange(AP.BioregTop5.BA.trees, AP.BioregTop5.VCF.grass, ncol=2)
```

Subsetting by Taxonomy and Height

```
# Subsetting by Taxonomy and Height
# *******************************

# Compute Single Cover Values Tables
# ==================================
# Trees (by default my.growth_forms=c("Tree/Palm", "Tree Mallee"))
veg.cover.trees.gt5 = single_cover_value(AP.BioregTop5.l$veg.PI, by.growth_form=TRUE, my.grow
th_forms=c("Tree/Palm", "Tree Mallee"), min.height=5)

# These are the default values for the arguments of the function 'single_cover_value'
# So the following function call would produce exactly the same results
veg.cover.trees.gt5.2 = single_cover_value(AP.BioregTop5.l$veg.PI)
all.equal(veg.cover.trees.gt5, veg.cover.trees.gt5.2)
```

```
## [1] TRUE
```

```
rm(veg.cover.trees.gt5.2)


# Enrich DF
# =========

# Add: Bioregion, Longitude, Latitude
# ---------------------------------

# Both DF have different number of rows
dim(veg.cover.trees.gt5)
```

```
## [1] 225    2
```

```
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231   44
```

```
# Enrich with: Bioregion, Latitude, and Longitude
AP.BioregTop5.VCF.trees.gt5.df  = merge(veg.cover.trees.gt5, AP.BioregTop5.l$site.info, by="s
ite_unique")[,c(names(veg.cover.trees.gt5), "bioregion.f", "longitude", "latitude")]
AP.BioregTop5.VCF.trees.gt5.df = na.omit(AP.BioregTop5.VCF.trees.gt5.df)
head(AP.BioregTop5.VCF.trees.gt5.df)
```

```
##         site_unique percentCover bioregion.f longitude  latitude
## 1 NSAMDD0001-56965          0.00         MDD   142.5602 -34.18392
## 2 NSAMDD0002-56952          3.56         MDD   142.6026 -34.20482
## 3 NSAMDD0003-56968          0.00         MDD   142.6041 -34.20754
## 4 NSAMDD0004-56953          8.51         MDD   142.5594 -34.16537
## 5 NSAMDD0005-56969          6.93         MDD   143.2039 -33.65619
## 6 NSAMDD0006-56954          8.22         MDD   143.1665 -33.66432
```

```
summary(AP.BioregTop5.VCF.trees.gt5.df)
```

```
##             site_unique    percentCover     bioregion.f    longitude
##   NSAMDD0001-56965:  1   Min.   : 0.000   FLB:48       Min.   :132.9
##   NSAMDD0002-56952:  1   1st Qu.: 0.000   GFU:41       1st Qu.:136.3
##   NSAMDD0003-56968:  1   Median : 0.200   MDD:50       Median :138.2
##   NSAMDD0004-56953:  1   Mean   : 6.478   SSD:46       Mean   :138.3
##   NSAMDD0005-56969:  1   3rd Qu.: 7.720   STP:40       3rd Qu.:139.3
##   NSAMDD0006-56954:  1   Max.   :70.000                Max.   :145.8
##   (Other)         :219
##      latitude
##   Min.   :-35.08
##   1st Qu.:-33.09
##   Median :-28.43
##   Mean   :-27.35
##   3rd Qu.:-23.83
##   Max.   :-14.05
##
```

```
names(AP.BioregTop5.VCF.trees.gt5.df)
```

```
## [1] "site_unique"  "percentCover" "bioregion.f"  "longitude"
## [5] "latitude"
```

```
# Graphical Visualisation
# =======================

# Map with circle size = Percent Cover
# -----------------------------------
AP.BioregTop5.VCF.trees.gt5.p1 =
ggplot(data=AP.BioregTop5.VCF.trees.gt5.df, aes(x=longitude, y=latitude, colour=bioregion.f,
 fill=bioregion.f), alpha =0.5) +
geom_point(aes(size=percentCover), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# Boxplot
# -------
AP.BioregTop5.VCF.trees.gt5.p2 =
ggplot(AP.BioregTop5.VCF.trees.gt5.df, aes(x=bioregion.f, y=percentCover, color=bioregion.f))
+ geom_boxplot() +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(name="Bioregions (Top 5)", values = c("darkgreen", "magenta", "red", "yello
w", "cyan")) +
labs(x="Bioregion (5 most sampled)", y = "Trees Cover Fraction (>= 5m)") +
theme(plot.title = element_text(hjust = 0.5))

# Plot both graphs
# ----------------
grid.arrange(AP.BioregTop5.VCF.trees.gt5.p1, AP.BioregTop5.VCF.trees.gt5.p2, nrow=2)
```
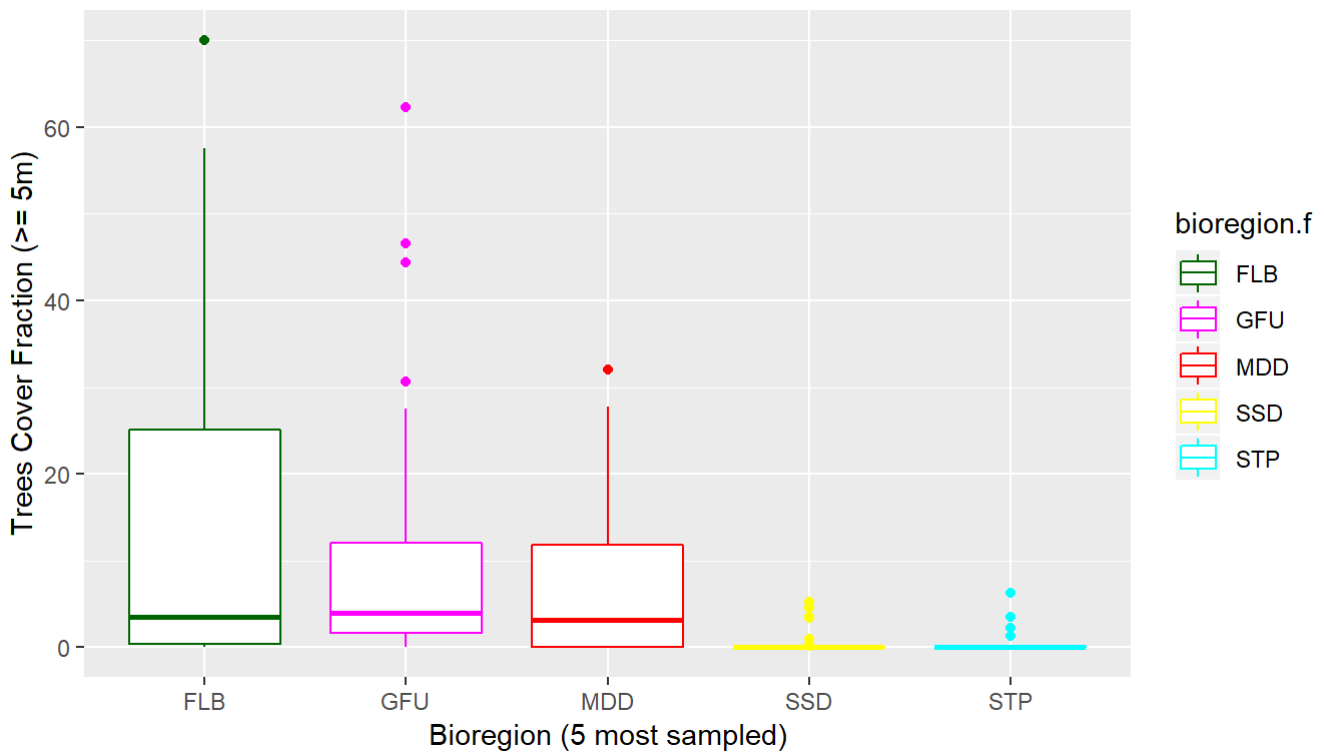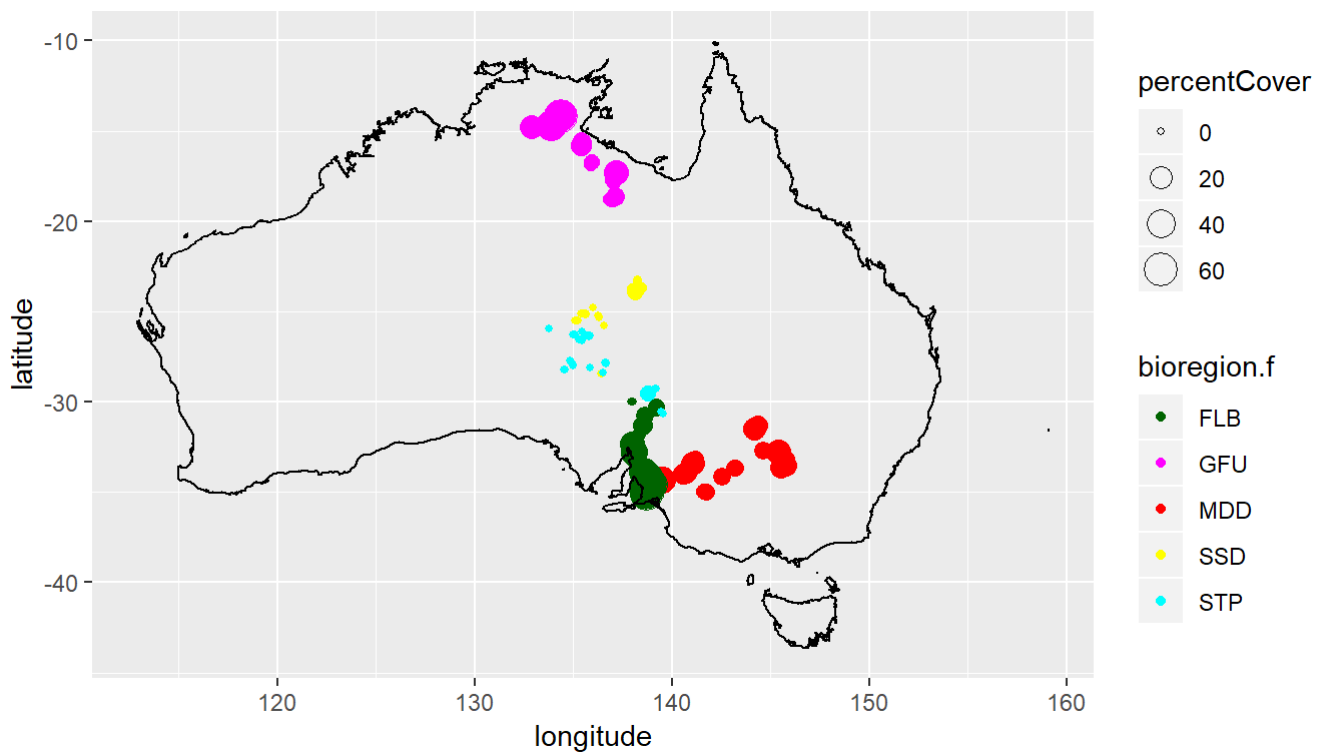
```
#grid.arrange(AP.BioregTop5.VCF.trees.gt5.p1, AP.BioregTop5.VCF.trees.gt5.p2, ncol=2)
```

BASAL AREA (OR NUMBER OF BASAL WEDGE HITS): `basal_area` function (for 5 most sampled bioregions).

The `basal_area` function in the `ausplotsR` package calculates the Basal Area or Number of Basal Wedge Hits of plots using the raw basal wedge data returned by the `get_ausplots` function also in the `ausplotsR` package.

The `basal_area` function has the following arguments: * `veg.basal` : Name of the data frame containing the Raw Basal Wedge Data generated by the `get_ausplots` function in the `ausplotsR` package. * `by.spp` : Whether Basal Wedge values are computed by Plot or Individual Species. This argument can take two values: * `FALSE` : The function returns combined values per plot. This is the default value for the argument. * `TRUE` : Returns values for individual

species
* `by.hits` : Whether the Basal Wedge values are computed as Basal Area or Number of Individual Basal Wedge Hits. This argument can take two values: * `FALSE` : The function returns the Basal Area (in m2/ha). This is the default value for the argument. * `TRUE` : Returns the Number of Individual Basal Wedge Hits.

This function returns a data frame with two columns. In this data frame rows correspond to representing Plots (or species by plots), the first column contains the unique sites names and the second column contains the Basal Area or Hit Scores depending on the value assigned to the `by.hits` argument.

In this section we will:

- Compute the Basal Area for each plot (m2/ha) using the `basal_area` function.
- Enrich the data frame containing the Basal Area data with additional information (i.e. plot -Site-Visit-, bioregion, longitude, and latitude).
- Display Basal Areas on map of Australia (with Dots size proportional to Basal Area).
- Boxplot of Basal Areas by Bioregion.

```
# Calculate Basal Area
# ====================
AP.BioregTop5.BA = basal_area(AP.BioregTop5.l$veg.basal)
summary(AP.BioregTop5.BA)
```

```
##   site_unique        basal_area_m2_ha
##   Length:115         Min.   : 0.2857
##   Class :character   1st Qu.: 1.8424
##   Mode  :character   Median : 3.9889
##                      Mean   : 5.9276
##                      3rd Qu.: 6.9861
##                      Max.   :40.2500
```

```
head(AP.BioregTop5.BA)
```

```
##         site_unique basal_area_m2_ha
## 1 NSAMDD0002-56952          4.583333
## 2 NSAMDD0004-56953          4.805556
## 3 NSAMDD0005-56969          5.538889
## 4 NSAMDD0006-56954          4.077778
## 5 NSAMDD0007-56970          4.205556
## 6 NSAMDD0009-56971         14.694444
```

```
# Enrich DF
# =========
# Preparation
colnames(AP.BioregTop5.BA)
```

```
## [1] "site_unique"      "basal_area_m2_ha"
```

```
summary(AP.BioregTop5.BA)
```

```
##   site_unique        basal_area_m2_ha
##  Length:115          Min.    : 0.2857
##  Class :character     1st Qu.: 1.8424
##  Mode  :character     Median : 3.9889
##                       Mean    : 5.9276
##                       3rd Qu.: 6.9861
##                       Max.    :40.2500
```

```
head(AP.BioregTop5.BA)
```

```
##           site_unique basal_area_m2_ha
## 1 NSAMDD0002-56952          4.583333
## 2 NSAMDD0004-56953          4.805556
## 3 NSAMDD0005-56969          5.538889
## 4 NSAMDD0006-56954          4.077778
## 5 NSAMDD0007-56970          4.205556
## 6 NSAMDD0009-56971         14.694444
```

```
# Add: Bioregion, Longitude, Latitude
# ----------------------------------
# Both DF have different number of rows
dim(AP.BioregTop5.BA)
```

```
## [1] 115    2
```

```
dim(AP.BioregTop5.l$site.info)
```

```
## [1] 231   44
```

```
# Enrich with: Bioregion, Latitude, and Longitude
AP.BioregTop5.BA  = merge(AP.BioregTop5.BA, AP.BioregTop5.l$site.info, by="site_unique")[,c(n
ames(AP.BioregTop5.BA), "bioregion.f", "longitude", "latitude")]
AP.BioregTop5.BA = na.omit(AP.BioregTop5.BA)
head(AP.BioregTop5.BA)
```

```
##           site_unique basal_area_m2_ha bioregion.f longitude   latitude
## 1 NSAMDD0002-56952          4.583333          MDD  142.6026 -34.20482
## 2 NSAMDD0004-56953          4.805556          MDD  142.5594 -34.16537
## 3 NSAMDD0005-56969          5.538889          MDD  143.2039 -33.65619
## 4 NSAMDD0006-56954          4.077778          MDD  143.1665 -33.66432
## 5 NSAMDD0007-56970          4.205556          MDD  141.1608 -33.37998
## 6 NSAMDD0009-56971         14.694444          MDD  141.0655 -33.44049
```

```
summary(AP.BioregTop5.BA)
```

```
##   site_unique        basal_area_m2_ha  bioregion.f   longitude
##   Length:115         Min.   : 0.2857   FLB:38        Min.   :132.9
##   Class :character   1st Qu.: 1.8424   GFU:34        1st Qu.:137.0
##   Mode  :character   Median : 3.9889   MDD:28        Median :138.6
##                      Mean   : 5.9276   SSD: 9        Mean   :138.6
##                      3rd Qu.: 6.9861   STP: 6        3rd Qu.:139.4
##                      Max.   :40.2500                 Max.   :145.8
##      latitude
##   Min.   :-35.08
##   1st Qu.:-33.35
##   Median :-30.41
##   Mean   :-27.05
##   3rd Qu.:-18.27
##   Max.   :-14.05
```

```r
names(AP.BioregTop5.BA)
```

```
## [1] "site_unique"     "basal_area_m2_ha" "bioregion.f"
## [4] "longitude"       "latitude"
```

```r
# Graphical Visualisation
# =======================

# Map with circle size = Basal Area (m2/ha)
# -----------------------------------------
AP.BioregTop5.BA.p1 =
ggplot(data=AP.BioregTop5.BA, aes(x=longitude, y=latitude, colour=bioregion.f, fill=bioregio
n.f), alpha =0.5) +
geom_point(aes(size=basal_area_m2_ha), pch=21) +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
geom_polygon(data=fortify(aus.sp), aes(x=long, y=lat, group=group), col="black", fill=NA)

# Boxplot
# -------
AP.BioregTop5.BA.p2 =
ggplot(AP.BioregTop5.BA, aes(x=bioregion.f, y=basal_area_m2_ha, color=bioregion.f)) +
geom_boxplot() +
scale_colour_manual(values = c("darkgreen", "magenta", "red", "yellow", "cyan")) +
scale_fill_manual(name="Bioregions (Top 5)", values = c("darkgreen", "magenta", "red", "yello
w", "cyan")) +
labs(title="Basal Area per Bioregion",x="Bioregion (5 most sampled)", y = "Basal area (m^2/H
a)") +
theme(plot.title = element_text(hjust = 0.5))

# Plot both graphs
# ----------------
grid.arrange(AP.BioregTop5.BA.p1, AP.BioregTop5.BA.p2, nrow=2)
```
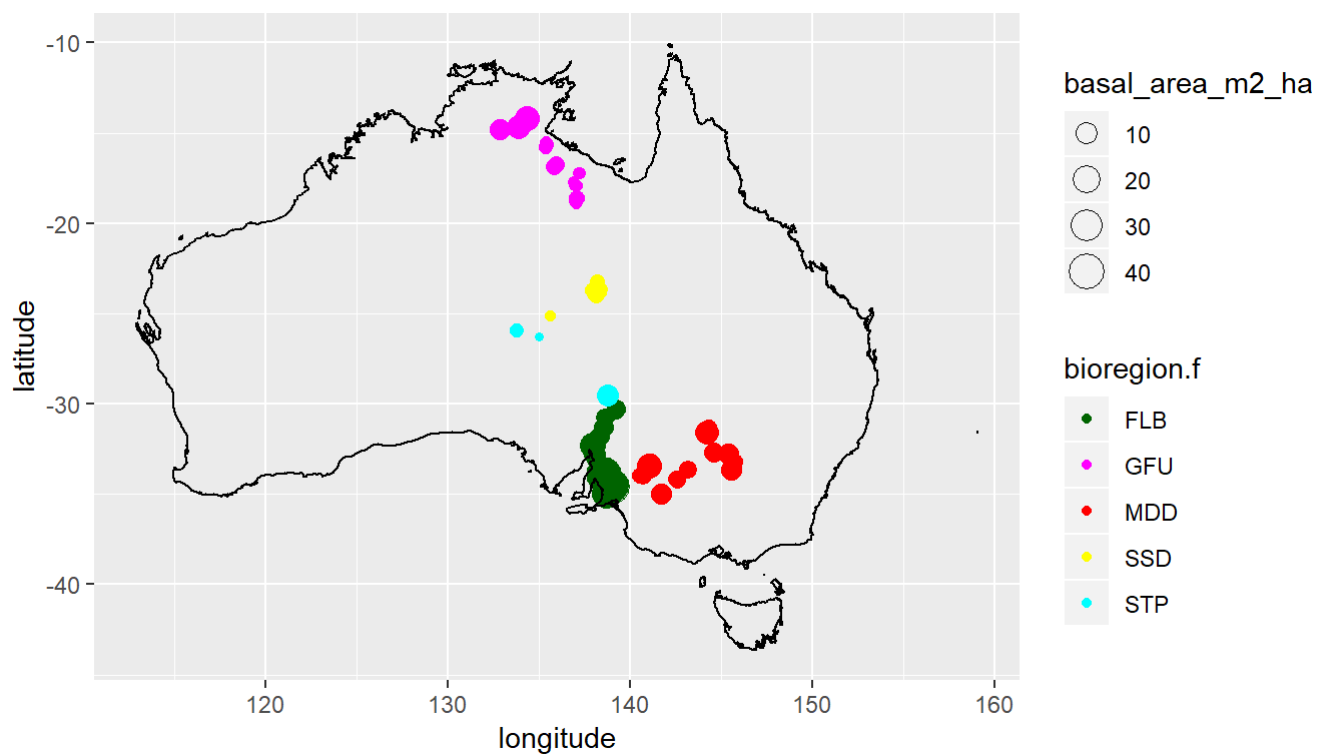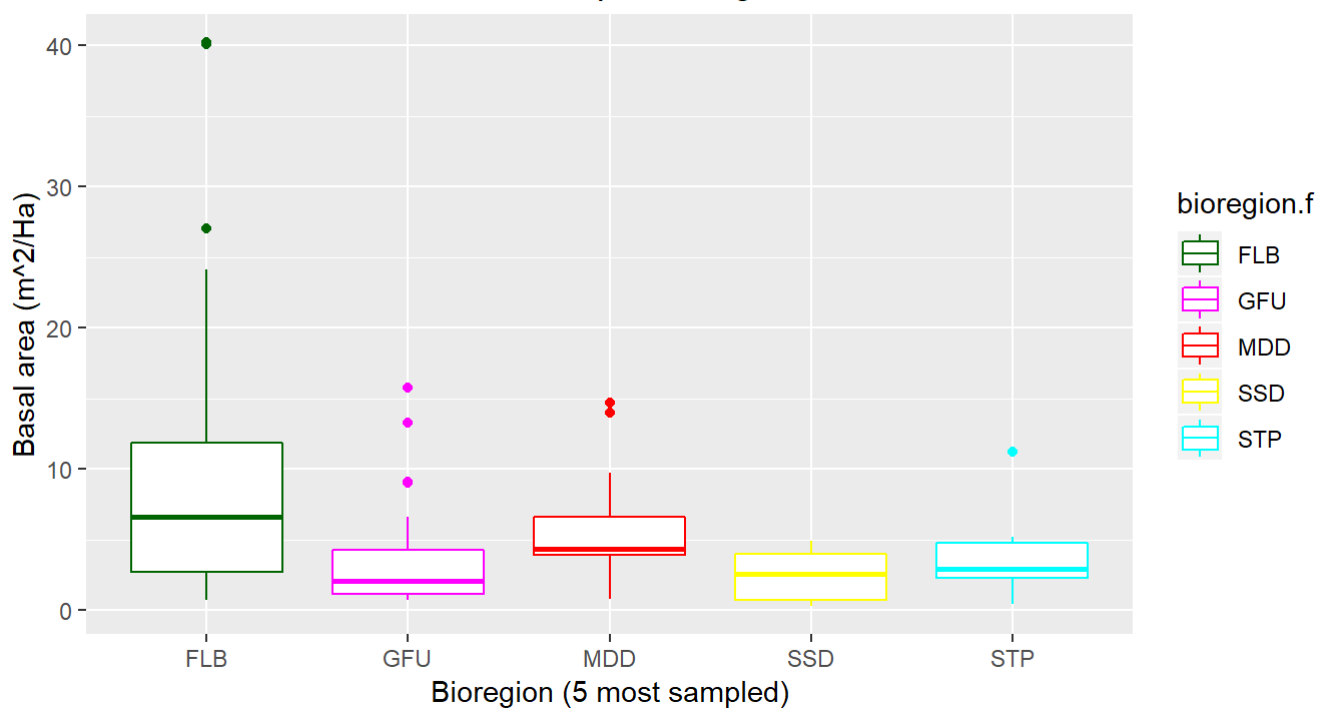
Basal Area per Bioregion

```
#grid.arrange(AP.BioregTop5.BA.p1, AP.BioregTop5.BA.p2, ncol=2)
```