ESA18 TERN Workshop: Ecosystem Surveillance (AusPlots)

"Understanding and using the ‘ausplotsR’ package and AusPlots data"

This document contains the Ecosystem Surveillance (AusPlots) section of TERN’s workshop at the Ecological Society of Australia 2018 conference.In this workshop we will explore the use of 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.

1. MANIPULATING AusPlots DATA:

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

1. MAP THE SITES

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

1. 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 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) .

1. PROPORTIONAL VEGETATION COVER (= FRACTIONAL COVER): `fractional\_cover’ function.

* Latitudinal pattern in proportional vegetation cover (for a random subset of 200 sites).

1. 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.

1. 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.

ACCESSING AND INSTALLING THE ausplotsR PACKAGE (plus its Dependencies)

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

## Install directly from github using the 'devtools' package  
## Thus, 'devtools' must be previouly installed  
#install.packages("devtools", repos="https://cloud.r-project.org/")  
#library(devtools)  
#install\_github("GregGuerin/ausplotsR", build\_vignettes = TRUE)  
  
## Load the package  
#library(ausplotsR)  
  
## Obtaining Help and Initial Steps  
#help(ausplotsR)  
#browseVignettes(package="ausplotsR")

Perhaps you also need to install the other packages required for this section of the workshop. A list of Compresenhive R Archive Network (CRAN) mirror URLs can be found [here](https://cran.r-project.org/mirrors.html).

## Select the repository (i.e. CRAN mirror URL)  
#my.repos = "https://cloud.r-project.org/"  
#my.repos = "https://cran.csiro.au/"  
  
## Install other required libraries  
#install.packages(c("ausplotsR", "vegan", "goeveg", "maps", "maptools", "mapdata", "sp", ggplot2", "gridExtra", "ggspatial", "dendextend"), repos=my.repos)

Before proceeding any further we load the other libraries that we will need for the workshop

# 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)

OBTAIN & EXPLORE AusPlots DATA: get\_ausplots function

This function extracts and compiles AusPlots data.

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.

In addition AusPlot data can be subset via the get\_ausplots function arguments in two other 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)).

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’,…). These data frames are interrelated and they all have a common variable, site\_location\_name. 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 retrieved (e.g. via my.Plot\_IDs or bounding\_box).

# Example 1: All available data 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 41 data.frame list  
## struct.summ 13 data.frame list  
## soil.subsites 10 data.frame list  
## soil.bulk 13 data.frame list  
## soil.char 32 data.frame list  
## veg.basal 10 data.frame list  
## veg.vouch 12 data.frame list  
## veg.PI 13 data.frame list

str(AP.data)

## List of 8  
## $ site.info :'data.frame': 3 obs. of 41 variables:  
## ..$ site\_location\_name : chr [1:3] "NTASTU0002" "QDAMGD0022" "SATFLB0004"  
## ..$ established\_date : chr [1:3] "2016-05-01T16:58:00" "2013-06-04T00:00:00" "2012-09-18T00:00:00"  
## ..$ description : chr [1:3] "Maryfield Station, 7.6km north north west of homestead. 27.5km south east of Larrimah" "Mackunda Downs Station, 500m east of homestead. 26km west of Middleton." "Brachina Gorge Heysen Range Lower. 63km North North East of Adelaide"  
## ..$ bioregion\_name : chr [1:3] "STU" "MGD" "FLB"  
## ..$ landform\_pattern : chr [1:3] "PLA" "ALP" "MOU"  
## ..$ landform\_element : chr [1:3] "PLA" "PLA" "HSL"  
## ..$ site\_slope : chr [1:3] "0" "1" "18"  
## ..$ site\_aspect : chr [1:3] NA "180" "135"  
## ..$ comments : chr [1:3] "Plot is flat. Low mound ( Likely anthropogenic) made up of ironstone gravels at the north west corner. Minimal "| \_\_truncated\_\_ "Astrebla pectinata / Cenchrus ciliaris / Astrebla elymoides low open tussock grassland on alluvial plain adjoin"| \_\_truncated\_\_ "Grazing impact high- goat tracks and droppings. Rabbit droppings also. Lots of Yellow footed rock wallabies clo"| \_\_truncated\_\_  
## ..$ outcrop\_lithology : chr [1:3] "NA" "NA" "NC"  
## ..$ other\_outcrop\_lithology : chr [1:3] "NC" "NA" "NC"  
## ..$ plot\_dimensions : chr [1:3] NA "100m x 100m." NA  
## ..$ site\_location\_visit\_id : int [1:3] 58429 53501 53705  
## ..$ visit\_start\_date : chr [1:3] "2016-05-01T16:58:00" "2013-05-18T09:34:00" "2012-09-18T00:00:00"  
## ..$ visit\_end\_date : chr [1:3] "2016-05-01T16:58:00" "2013-05-18T09:34:00" "2012-09-18T00:00:00"  
## ..$ visit\_notes : chr [1:3] "Corymbia polycarpa and Corymbia terminalis combined for Basal area\r\n\r\nunknown substrate in point intercept "| \_\_truncated\_\_ "" NA  
## ..$ location\_description : chr [1:3] "Maryfield Station, 7.6km north north west of homestead. 27.5km south east of Larimah" "Mackunda Station, north of Middleton." "Brachina Gorge Heysen Range Lower. 63km North North East of Adelaide"  
## ..$ erosion\_type : chr [1:3] "n/a" "G" "NC"  
## ..$ erosion\_abundance : chr [1:3] "X" "2" "NC"  
## ..$ erosion\_state : chr [1:3] "n/a" "NC" "NC"  
## ..$ microrelief : chr [1:3] "TM" "Z" "NC"  
## ..$ drainage\_type : int [1:3] 4 4 7  
## ..$ disturbance : chr [1:3] "0" "1L" "NC"  
## ..$ climatic\_condition : chr [1:3] "DRY" "DRY" "Wet"  
## ..$ vegetation\_condition : chr [1:3] "AVG" "DRY" "Active vegetative growth"  
## ..$ observer\_veg : int [1:3] 1 3 16  
## ..$ observer\_soil : int [1:3] 2 2 1  
## ..$ described\_by : int [1:3] 1 3 16  
## ..$ pit\_marker\_easting : int [1:3] 326265 529568 839490  
## ..$ pit\_marker\_northing : int [1:3] 8256078 7526350 6528576  
## ..$ pit\_marker\_mga\_zones : int [1:3] 53 54 53  
## ..$ pit\_marker\_datum : chr [1:3] "WGS84" "WGS84" "GDA94"  
## ..$ pit\_marker\_location\_method: chr [1:3] NA "GPS" "GPS"  
## ..$ soil\_observation\_type : chr [1:3] "P" "P" "P"  
## ..$ a\_s\_c : chr [1:3] NA "NC" NA  
## ..$ plot\_is\_100m\_by\_100m : logi [1:3] TRUE TRUE TRUE  
## ..$ plot\_is\_aligned\_to\_grid : logi [1:3] TRUE TRUE TRUE  
## ..$ plot\_is\_permanently\_marked: logi [1:3] TRUE TRUE TRUE  
## ..$ latitude : num [1:3] -15.8 -22.4 -31.3  
## ..$ longitude : num [1:3] 133 141 139  
## ..$ point : chr [1:3] "SW" "SW" "SW"  
## $ struct.summ :'data.frame': 3 obs. of 13 variables:  
## ..$ site\_location\_name : chr [1:3] "QDAMGD0022" "SATFLB0004" "NTASTU0002"  
## ..$ phenology\_comment : chr [1:3] "" "Ptilotus obovatus var. obovatus flowering. Triodia sp. has no seeds. No fruit on Callitris glaucophylla - no e"| \_\_truncated\_\_ "NC"  
## ..$ upper\_1\_dominant : chr [1:3] "" "SAT 000251" "NTA017194"  
## ..$ upper\_2\_dominant : chr [1:3] "" "SAT 000229" "NTA017232"  
## ..$ upper\_3\_dominant : chr [1:3] "" NA "NTA017084"  
## ..$ mid\_1\_dominant : chr [1:3] "" "SAT 000244" NA  
## ..$ mid\_2\_dominant : chr [1:3] "" "SAT 000261" NA  
## ..$ mid\_3\_dominant : chr [1:3] "" NA NA  
## ..$ ground\_1\_dominant : chr [1:3] "QDA 003325" "SAT 000233" "NTA017070"  
## ..$ ground\_2\_dominant : chr [1:3] "QDA 003293" NA "NTA017076"  
## ..$ ground\_3\_dominant : chr [1:3] "QDA 003325" NA "NTA017082"  
## ..$ description : chr [1:3] "Astrebla pectinata / Cenchrus ciliaris / Astrebla elymoides low open tussock grassland with scattered ." "Callitris glaucophylla / Eucalyptus intertexta low woodland. A mid-stratum dominated by Rhagodia paradoxa and H"| \_\_truncated\_\_ "Corymbia terminalis mixed mid woodland with Corymbia polycarpa / Eucalyptus pruinosa/ Eucalyptus chlorophylla "| \_\_truncated\_\_  
## ..$ mass\_flowering\_event: logi [1:3] FALSE FALSE FALSE  
## $ soil.subsites:'data.frame': 27 obs. of 10 variables:  
## ..$ site\_location\_name : chr [1:27] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" ...  
## ..$ subsite\_id : chr [1:27] "1" "2" "3" "4" ...  
## ..$ zone : int [1:27] 54 54 54 54 54 54 54 54 54 53 ...  
## ..$ easting : int [1:27] 529581 529580 529582 529592 529595 529620 529664 529587 529663 268530 ...  
## ..$ northing : int [1:27] 7526343 7526338 7526335 7526338 7526345 7526378 7526417 7526423 7526333 6531529 ...  
## ..$ ten\_to\_twenty\_barcode : chr [1:27] "QDA 051589" "QDA 051592" "QDA 051595" "QDA 051598" ...  
## ..$ zero\_to\_ten\_barcode : chr [1:27] "QDA 051588" "QDA 051591" "QDA 051594" "QDA 051597" ...  
## ..$ twenty\_to\_thirty\_barcode: chr [1:27] "QDA 051590" "QDA 051593" "QDA 051596" "QDA 051599" ...  
## ..$ comments : chr [1:27] "bare ground" "between grass tussocks" "between grass tussocks" "bare ground" ...  
## ..$ metagenomic\_barcode : chr [1:27] "QDA 053721" "QDA 053722" "QDA 053723" "QDA 053724" ...  
## $ soil.bulk :'data.frame': 6 obs. of 13 variables:  
## ..$ site\_location\_name : chr [1:6] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "NTASTU0002" ...  
## ..$ sample\_id : chr [1:6] "0" "1" "2" "0" ...  
## ..$ paper\_bag\_weight : logi [1:6] NA NA NA NA NA NA  
## ..$ oven\_dried\_weight\_in\_bag: logi [1:6] NA NA NA NA NA NA  
## ..$ ring\_weight : logi [1:6] NA NA NA NA NA NA  
## ..$ gravel\_weight : int [1:6] NA NA NA 0 0 0  
## ..$ ring\_volume : num [1:6] NA NA NA 209 209 ...  
## ..$ gravel\_volume : int [1:6] NA NA NA 0 0 0  
## ..$ fine\_earth\_weight\_in\_bag: int [1:6] NA NA NA 0 0 0  
## ..$ fine\_earth\_weight : int [1:6] NA NA NA 0 0 0  
## ..$ fine\_earth\_volume : num [1:6] NA NA NA 209 209 ...  
## ..$ fine\_earth\_bulk\_density : int [1:6] NA NA NA 0 0 0  
## ..$ gravel\_bulk\_density : int [1:6] NA NA NA 0 0 0  
## $ soil.char :'data.frame': 21 obs. of 32 variables:  
## ..$ site\_location\_name : chr [1:21] "SATFLB0004" "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" ...  
## ..$ upper\_depth : num [1:21] 0 0.5 0.8 0.6 0.7 0.1 0.3 0.6 0.2 0 ...  
## ..$ lower\_depth : num [1:21] 0.1 0.6 0.9 0.7 0.8 0.2 0.4 0.7 0.3 0.1 ...  
## ..$ horizon : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ texture\_grade : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ texture\_qualifier : chr [1:21] NA "NC" "NC" "NC" ...  
## ..$ texture\_modifier : chr [1:21] NA "NC" "NC" "NC" ...  
## ..$ colour\_when\_moist : chr [1:21] NA "NC" "NC" "NC" ...  
## ..$ colour\_when\_dry : chr [1:21] NA "NC" "NC" "NC" ...  
## ..$ mottles\_colour : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ mottles\_abundance : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ mottles\_size : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ segregations\_abundance: chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ segregations\_size : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ segregations\_nature : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ segregations\_form : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ comments : chr [1:21] NA NA NA NA ...  
## ..$ collected\_by : int [1:21] 8 4 4 4 4 NA 4 NA 4 4 ...  
## ..$ smallest\_size\_1 : chr [1:21] "11" "11" "11" "11" ...  
## ..$ smallest\_size\_2 : logi [1:21] NA NA NA NA NA NA ...  
## ..$ effervescence : chr [1:21] "N" "N" "N" "N" ...  
## ..$ ec : num [1:21] 0.07 0.18 0.54 0.38 0.52 0.03 0.07 1.84 0.06 0.04 ...  
## ..$ ph : num [1:21] 6.3 8.3 8.1 8.3 8.1 7 8 8.3 7.9 7.8 ...  
## ..$ pedality\_grade : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ pedality\_fabric : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ next\_size\_type\_2 : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ next\_size\_type\_1 : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ smallest\_size\_type\_2 : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ smallest\_size\_type\_1 : chr [1:21] "NC" "NC" "NC" "NC" ...  
## ..$ next\_size\_2 : logi [1:21] NA NA NA NA NA NA ...  
## ..$ next\_size\_1 : chr [1:21] "11" "11" "11" "11" ...  
## ..$ layer\_barcode : chr [1:21] "SAT005230" "QDA 051583" "QDA 051586" "QDA 051584" ...  
## $ veg.basal :'data.frame': 70 obs. of 10 variables:  
## ..$ site\_location\_name : chr [1:70] "NTASTU0002" "SATFLB0004" "SATFLB0004" "SATFLB0004" ...  
## ..$ site\_location\_visit\_id : int [1:70] 58429 53705 53705 53705 53705 53705 53705 53705 53705 53705 ...  
## ..$ site\_location\_id : int [1:70] 61138 60122 60122 60122 60122 60122 60122 60122 60122 60122 ...  
## ..$ point\_id : chr [1:70] "NE" "W" "SW" "SE" ...  
## ..$ herbarium\_determination: chr [1:70] "Dead Tree/Shrub" "Alectryon oleifolius" "Alectryon oleifolius" "Alectryon oleifolius" ...  
## ..$ veg\_barcode : chr [1:70] "NO\_BARCODE\_DEAD\_TREE\_804159" "SAT 000242" "SAT 000242" "SAT 000242" ...  
## ..$ hits : int [1:70] 1 1 3 4 1 1 1 20 17 6 ...  
## ..$ basal\_area\_factor : num [1:70] 0.1 0.25 0.5 0.5 0.5 0.25 0.5 0.25 0.5 0.5 ...  
## ..$ basal\_area : num [1:70] 0.1 0.25 1.5 2 0.5 0.25 0.5 5 8.5 3 ...  
## ..$ site\_unique : chr [1:70] "NTASTU0002-58429" "SATFLB0004-53705" "SATFLB0004-53705" "SATFLB0004-53705" ...  
## $ veg.vouch :'data.frame': 149 obs. of 12 variables:  
## ..$ site\_location\_name : chr [1:149] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" ...  
## ..$ veg\_barcode : chr [1:149] "QDA 003331" "NO\_BARCODE\_FORB\_950413164" "NO\_BARCODE\_GRASS\_656236361" "NO\_BARCODE\_DEAD\_TREE\_558409020" ...  
## ..$ herbarium\_determination : chr [1:149] "Glinus lotoides" "Annual forb" "Annual grass" "Dead tree/shrub" ...  
## ..$ is\_uncertain\_determination: logi [1:149] FALSE NA NA NA NA NA ...  
## ..$ visit\_start\_date : chr [1:149] "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 [1:149] 53501 53501 53501 53501 53705 53705 53705 53705 58429 58429 ...  
## ..$ primary\_gen\_barcode : chr [1:149] "QDA 003332" NA NA NA ...  
## ..$ secondary\_gen\_barcode\_1 : chr [1:149] NA NA NA NA ...  
## ..$ secondary\_gen\_barcode\_2 : chr [1:149] NA NA NA NA ...  
## ..$ secondary\_gen\_barcode\_3 : chr [1:149] NA NA NA NA ...  
## ..$ secondary\_gen\_barcode\_4 : chr [1:149] NA NA NA NA ...  
## ..$ site\_unique : chr [1:149] "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD0022-53501" ...  
## $ veg.PI :'data.frame': 3217 obs. of 13 variables:  
## ..$ site\_location\_name : chr [1:3217] "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" "QDAMGD0022" ...  
## ..$ site\_location\_visit\_id : int [1:3217] 53501 53501 53501 53501 53501 53501 53501 53501 53501 53501 ...  
## ..$ transect : Factor w/ 15 levels "E2-W2","E4-W4",..: 13 13 14 14 14 14 14 14 14 14 ...  
## ..$ point\_number : int [1:3217] 96 27 6 7 8 9 10 11 12 13 ...  
## ..$ veg\_barcode : chr [1:3217] NA NA "QDA 003319" NA ...  
## ..$ herbarium\_determination: chr [1:3217] NA NA "Cenchrus ciliaris" NA ...  
## ..$ substrate : chr [1:3217] "Bare" "Bare" "Litter" "Litter" ...  
## ..$ in\_canopy\_sky : logi [1:3217] NA NA FALSE NA NA NA ...  
## ..$ dead : logi [1:3217] NA NA FALSE NA NA NA ...  
## ..$ growth\_form : chr [1:3217] NA NA "Tussock grass" NA ...  
## ..$ height : num [1:3217] NA NA 0.3 NA NA NA 0.1 NA NA NA ...  
## ..$ hits\_unique : chr [1:3217] "W1-E1 96" "W1-E1 27" "W3-E3 6" "W3-E3 7" ...  
## ..$ site\_unique : chr [1:3217] "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD0022-53501" "QDAMGD0022-53501" ...

# 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 41 data.frame list  
## veg.vouch 12 data.frame list  
## veg.PI 13 data.frame list

#str(AP.data) # Similar to Example 1 (can run uncommented if curious)   
  
  
# Example 3: 'site\_info', 'veg.PI', and 'basal.wedge' data for al 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 1.057 mins

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

## Length Class Mode  
## site.info 41 data.frame list  
## veg.basal 10 data.frame list  
## veg.PI 13 data.frame list

#str(AP.data) # Similar to Example 1 (can run uncommented if curious)   
  
# Explore 'site\_info' data  
dim(AP.data$site.info)

## [1] 607 41

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"

head(AP.data$site.info)

## site\_location\_name established\_date  
## 1 WAANUL0003 2014-07-01T15:18:42  
## 2 NTAFIN0007 2011-10-12T00:00:00  
## 3 SAASTP0003 2012-06-29T00:00:00  
## 4 WAAHAM0006 2014-06-28T17:00:42  
## 5 QDASSD0015 2014-05-08T00:00:00  
## 6 QDASSD0001 2014-04-30T00:00:00  
## description  
## 1 Plumridge Lakes Nature Reserve. About 178kms west south west of Tjuntjuntjara in Great Victoria Desert.  
## 2 Calcareous rise north of Waterhouse Range, Owen Springs Reserve. 52km South west of Alice Springs  
## 3 Witjira National Park. West of Blood Creek Bore. 184km South East of Kulgera. Plot approx.. 100m north of track.  
## 4 About 28kms south east of Cocklebiddy and 5kms north west of Eyre Bird Observatory.  
## 5 Cravens Peak Reserve. 56km north west of Carlo homestead  
## 6 Ethabuka Station. 10.7 km north of homestead  
## bioregion\_name landform\_pattern landform\_element site\_slope site\_aspect  
## 1 NUL PLA PLA 0 <NA>  
## 2 MAC PLT PLA 0 <NA>  
## 3 STP FLO PLA 1 90  
## 4 HAM DUN DUS 2 90  
## 5 CHC PLT PLA 0 <NA>  
## 6 SSD DUN DUC 4 75  
## comments  
## 1 Grazing effect is low- some evidence of camels with older broken trees. Very little surface strew. Site is flat plain. Introduced plant species- none noted. Homogeneity- site is in consistent patch of Acacia aneura / Acacia papyrocarpa woodland with the community extending 200m in all directions with just the odd emergent Allocasuarina pauper outside the plot.  
## 2 On limestone plateau. Grazing impact moderate- rabbit warrens within the site. Recent fire through the site though somewhat patchy in extent. Minimal impact of introduced plant species- -Cenchrus ciliaris present.  
## 3 Site covers areas of fine gravels with some light gibber patches with some silcrete stones. Site has less Chenopod cover than SAASTP0001 and 2, with less flood out influence. Heavier soils with high proportion of Sclerolaena cover. Some elevated areas. Minimal impact of introduced plant species- Malvastrum americanum var. americanum present.\r\n  
## 4 Very long unburnt- no scarring on any of the trees and no dead standing litter 30+ years. Grazing effect is low- some rabbits and kangaroos. Introduced plant effect- minimal. Dune slope in Dunefield. Wet- recent rain in previous day and also before that. Homogeneity- continues to the west 60m. Opens up to the north on the dune crest and another 50m to the east where it joins the road. Slope is 2 degree slope through the centre of the site to the north west but site is convex. Rises at the south east corner and the northern slope.  
## 5 Introduced plant effect low- none noted. Long unburnt- some dead litter through the centre of the plot but has been there for some time. Mix of cobbles and pebbles with cobbles dominating,. Mixed lithology- sandstone as well as some gibber. Homogeneity- Community opens up to the south. North east corner is where the track is. To the west it extends 70m and the same to the north. Large stand of sparse Acacia aneura shrubland.  
## 6 Dune slope- South west corner is just over the crest of the dune and North east corner is just bordering the a stand of Acacia georginae. Dune is running to the north east. Site is across the slope of the dune. Long unburnt- at least 20 years but probably longer. No grazing impact. Camels around site but no evidence within the site. Minimal rabbits. No introduced plant species seen. 6 degree slope, slightly uneven through the centre where it flattens out somewhat. Aspect is to the North East ( 45 degrees).  
## outcrop\_lithology other\_outcrop\_lithology plot\_dimensions  
## 1 S S <NA>  
## 2 KC QS 100 x 100 m.  
## 3 NC NC <NA>  
## 4 S S <NA>  
## 5 NA NA 100 x 100 m.  
## 6 NA NA 100m x 100m.  
## site\_location\_visit\_id visit\_start\_date visit\_end\_date  
## 1 56946 2014-07-01T15:18:42 2014-07-01T15:18:42  
## 2 58013 2016-03-02T00:00:00 2016-03-02T00:00:00  
## 3 53721 2012-06-29T00:00:00 2012-12-29T00:00:00  
## 4 56944 2014-06-28T17:00:42 2014-06-28T17:00:42  
## 5 56924 2014-05-08T00:00:00 2014-05-08T00:00:00  
## 6 53756 2014-04-30T00:00:00 2014-04-30T00:00:00  
## visit\_notes  
## 1 Basal wedge measure difficult due to shrubby nature of site.  
## 2 Revisit collected Point intercept, vouchered plant specimens,DNA and Metagenomic samples.   
## 3   
## 4 Most of the cryptogam is moss. \r\n\r\n\r\n  
## 5 \r\ntransect E3-W3 3 points 69 to 97 missing  
## 6 Points 53 onwards on north to south 3 missing\r\n\r\n\r\n\r\nnumbers on first and second east west transects need swapping around.  
## location\_description  
## 1 120kms west of Tjuntjuntjara in Great Victoria Desert. Myall and Mulga woodland  
## 2 Owen Springs, north of Waterhouse Range  
## 3 Witjira National Park. West of Blood Creek Bore. 184km South East of Kulgera  
## 4 28kms SE of Cocklebiddy and 5kms NW of Eyre Bird Observatory. Low Mallee woodland on sand.  
## 5 Cravens Peak reserve. 56km north west of Carlo homestead  
## 6 Ethabuka station. 10.7 km north of homestead  
## erosion\_type erosion\_abundance erosion\_state microrelief drainage\_type  
## 1 NC X NC X 4  
## 2 NC NC NC NC 7  
## 3 NC NC NC NC 7  
## 4 R 1 P Z 6  
## 5 S 1 S U 4  
## 6 n/a X n/a Z 6  
## disturbance climatic\_condition vegetation\_condition observer\_veg  
## 1 0 DRY DRY 1  
## 2 NC DRY AVG 1  
## 3 NC Dry Disturbed natural 5  
## 4 0 WET FFR 1  
## 5 0 DRY DRY 1  
## 6 0 DRY DRY 1  
## observer\_soil described\_by pit\_marker\_easting pit\_marker\_northing  
## 1 2 1 724008 6710139  
## 2 11 1 340939 7351405  
## 3 4 5 507285 7093980  
## 4 2 1 245824 6432085  
## 5 2 1 212761 7432990  
## 6 2 1 242808 7369711  
## pit\_marker\_mga\_zones pit\_marker\_datum pit\_marker\_location\_method  
## 1 51 WGS84 GPS  
## 2 53 WGS84 <NA>  
## 3 53 WGS84 DGPS  
## 4 52 WGS84 GPS  
## 5 54 WGS84 GPS  
## 6 54 WGS84 GPS  
## soil\_observation\_type a\_s\_c plot\_is\_100m\_by\_100m plot\_is\_aligned\_to\_grid  
## 1 P <NA> TRUE TRUE  
## 2 NC <NA> TRUE TRUE  
## 3 P <NA> TRUE TRUE  
## 4 P <NA> TRUE TRUE  
## 5 P <NA> TRUE TRUE  
## 6 P <NA> TRUE TRUE  
## plot\_is\_permanently\_marked latitude longitude point  
## 1 TRUE -29.71878 125.3158 SW  
## 2 TRUE -23.94119 133.4360 SW  
## 3 TRUE -26.27390 135.0730 SW  
## 4 TRUE -32.21914 126.3019 SW  
## 5 TRUE -23.18808 138.1931 SW  
## 6 TRUE -23.76407 138.4765 SW

# Explore 'veg\_PI' data  
dim(AP.data$veg.PI)

## [1] 660321 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 VCAAUA0001 58565 W1-E1 0  
## 2 VCAAUA0001 58565 W1-E1 1  
## 3 VCAAUA0001 58565 W1-E1 2  
## 4 VCAAUA0001 58565 W1-E1 2  
## 5 VCAAUA0001 58565 W1-E1 3  
## 6 VCAAUA0001 58565 W1-E1 4  
## veg\_barcode herbarium\_determination substrate  
## 1 VCA005003 Asterolasia trymalioides subsp. trymalioides Litter  
## 2 VCA005079 Celmisia pugioniformis Litter  
## 3 VCA005089 Poa hiemata Litter  
## 4 VCA005035 Leptorhynchos squamatus subsp. alpinus Litter  
## 5 VCA005089 Poa hiemata Litter  
## 6 VCA005079 Celmisia pugioniformis Litter  
## in\_canopy\_sky dead growth\_form height hits\_unique site\_unique  
## 1 FALSE FALSE Heath-shrub 0.05 W1-E1 0 VCAAUA0001-58565  
## 2 FALSE FALSE Forb 0.01 W1-E1 1 VCAAUA0001-58565  
## 3 FALSE FALSE Tussock grass 0.05 W1-E1 2 VCAAUA0001-58565  
## 4 FALSE FALSE Forb 0.03 W1-E1 2 VCAAUA0001-58565  
## 5 FALSE FALSE Tussock grass 0.03 W1-E1 3 VCAAUA0001-58565  
## 6 FALSE FALSE Forb 0.01 W1-E1 4 VCAAUA0001-58565

MANIPULATING AusPlots DATA

The retrieved data by the function ‘get\_ausplots’ can be manipulated as any other R data. However, the ‘deep’ structure of the data and interrelation of the data frames (via the common site\_loation\_name variable) can make manipulating the data a bit more daunting.

As an example, we will focus on the sites in the 5 most sampled Bioregions. We will first identify which are these regions, and then subset the sites in these regions.

#----------------------------------------------------------------------------------------  
# 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 SSD GFU STP PIL FLB MGD GUP BHC COO RIV MAC FIN NUL AUA CHC SYB EIU   
## 51 48 41 38 35 34 34 33 32 32 32 28 18 13 12 10 9 7   
## MUL BRT HAM MUR STU GVD AVW KAN SWA VIB ARP BBS CEK DAC DMR EYB GAW GES   
## 7 6 6 6 6 5 4 4 4 4 3 3 3 3 3 3 3 3   
## JAF LSD MAL PCK COP GAS MII DAB DAL ESP GSD NAN NSS   
## 3 3 3 3 2 2 2 1 1 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" "SSD" "GFU" "STP" "PIL"

#----------------------------------------------------------------------------------------  
# Subset data for the 5 most 'visited/sampled' Bioregions   
#----------------------------------------------------------------------------------------  
  
summary(AP.data)

## Length Class Mode  
## site.info 42 data.frame list  
## veg.basal 10 data.frame list  
## veg.PI 13 data.frame list

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

## [1] 607 42

AP.BioregTop5.l = AP.data  
AP.BioregTop5.l$site.info = AP.BioregTop5.l$site.info[AP.BioregTop5.l$site.info$bioregion\_name %in% Bioregs.Top5.s, ]  
dim(AP.BioregTop5.l$site.info)

## [1] 213 42

# 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" "DAB"  
## [12] "DAC" "DAL" "DMR" "EIU" "ESP" "EYB" "FIN" "FLB" "GAS" "GAW" "GES"  
## [23] "GFU" "GSD" "GUP" "GVD" "HAM" "JAF" "KAN" "LSD" "MAC" "MAL" "MDD"  
## [34] "MGD" "MII" "MUL" "MUR" "NAN" "NSS" "NUL" "PCK" "PIL" "RIV" "SSD"  
## [45] "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] "GFU" "MDD" "PIL" "SSD" "STP"

# Subset the 5 most sampled Bioregions in the 'veg.PI' data frame  
# ===============================================================  
dim(AP.BioregTop5.l$veg.PI)

## [1] 660321 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] 224160 13

# Subset the 5 most sampled Bioregions in the 'veg.basal' data frame  
# ==================================================================  
dim(AP.BioregTop5.l$veg.basal)

## [1] 7093 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] 1999 10

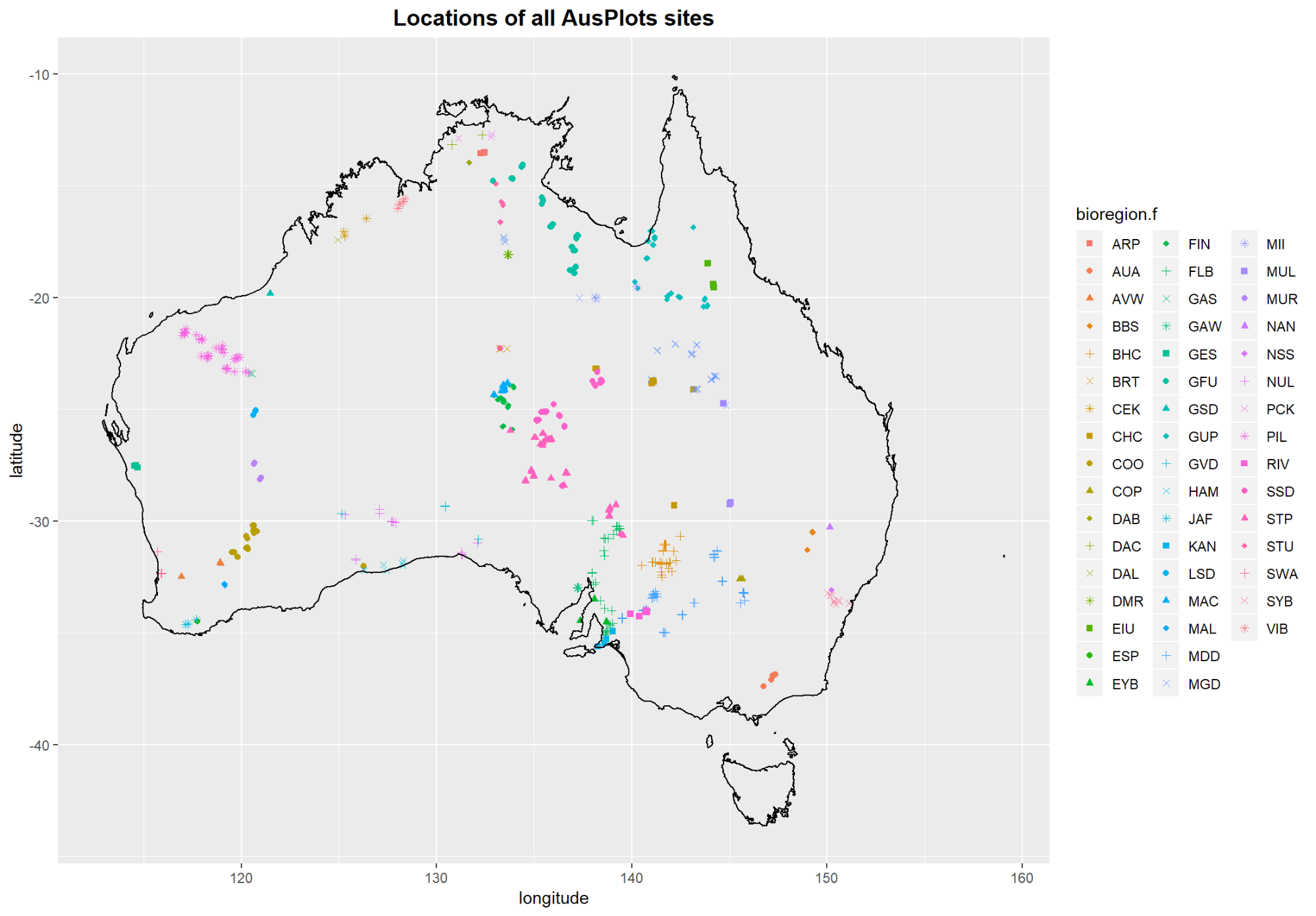
MAP THE SITES

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.

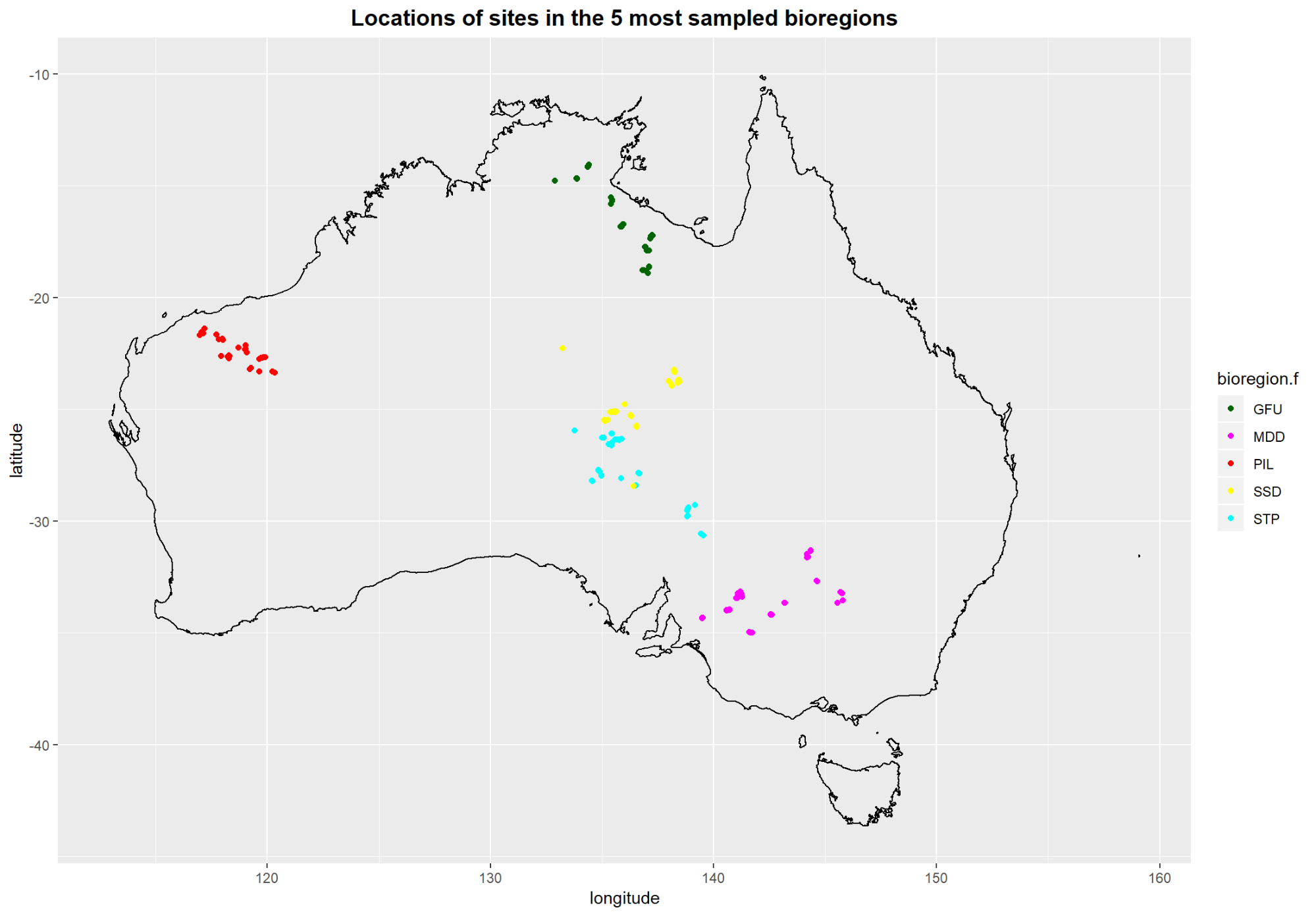
#----------------------------------------------------------------------------------------  
# Get and Prepare a Map of Australia  
#----------------------------------------------------------------------------------------  
  
# Maps in the package 'maps' are projected in longlat by default  
aus = map("worldHires", "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

aus.sp = map2SpatialPolygons(aus, IDs=aus$names, proj4string=CRS("+init=epsg:4326"))  
  
  
#----------------------------------------------------------------------------------------  
# Plot All AusPlots Sites on a Map of Australia  
#----------------------------------------------------------------------------------------  
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=rep(c(15:18,3:4,8),7)) + # 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)



#----------------------------------------------------------------------------------------  
# 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, fill=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)



SPECIES-LEVEL DATA: species\_table function and species occurence 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 occurence 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* (argument m\_kind = PA).
* *Percent Cover*: Based on total frequency of hits. This is the most commonly used metric (argument m\_kind = percent\_cover).
* *Frequency*: Based on proportional frequencies of presence on the 10 individual transects within a plot (argument m\_kind = freq). It can be a measure of importance for low cover species.
* *IVI*: A combination of cover and frequency (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 (argument cover\_type = PFC).
* *Opaque Canopy Cover* (*OCC*): Hits scored as ‘in canopy sky’ are retained (argument cover\_type = OCC).

# 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\_type="PFC")  
class(SppBYSites.BioregTop5)

## [1] "data.frame"

dim(SppBYSites.BioregTop5) # Number of rows and columns in the matrix: 574 Sites x 3024 Spp

## [1] 209 1091

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.oxycarpum Abutilon.sp.  
## NSAMDD0001-56965 0 0  
## NSAMDD0002-56952 0 0  
## NSAMDD0003-56968 0 0  
## NSAMDD0004-56953 0 0  
## NSAMDD0005-56969 0 0

# Enrich Table with: Site\_Location, Bioregion, Latitude, and Longitude  
# ====================================================================  
  
# Create a site\_visit variable in original dataset to relate both datasets  
# -------------------------------------------------------------------------  
AP.BioregTop5.l$site.info$Plot = paste(AP.BioregTop5.l$site.info$site\_location\_name,   
 AP.BioregTop5.l$site.info$site\_location\_visit\_id, sep="-")   
SppBYSites.BioregTop5$Plot = rownames(SppBYSites.BioregTop5)   
  
# Both DF have differente number of rows!  
dim(SppBYSites.BioregTop5)

## [1] 209 1092

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

## [1] 213 43

# Enrich with: Bioregion, Latitude, and Longitude   
# -----------------------------------------------  
SppBYSites.BioregTop5 = merge(SppBYSites.BioregTop5, AP.BioregTop5.l$site.info, by="Plot")[,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.oxycarpum" "Abutilon.sp." "Acacia.adoxa"

Species Abundance

In AusPlots data percent cover is used as a measure of 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

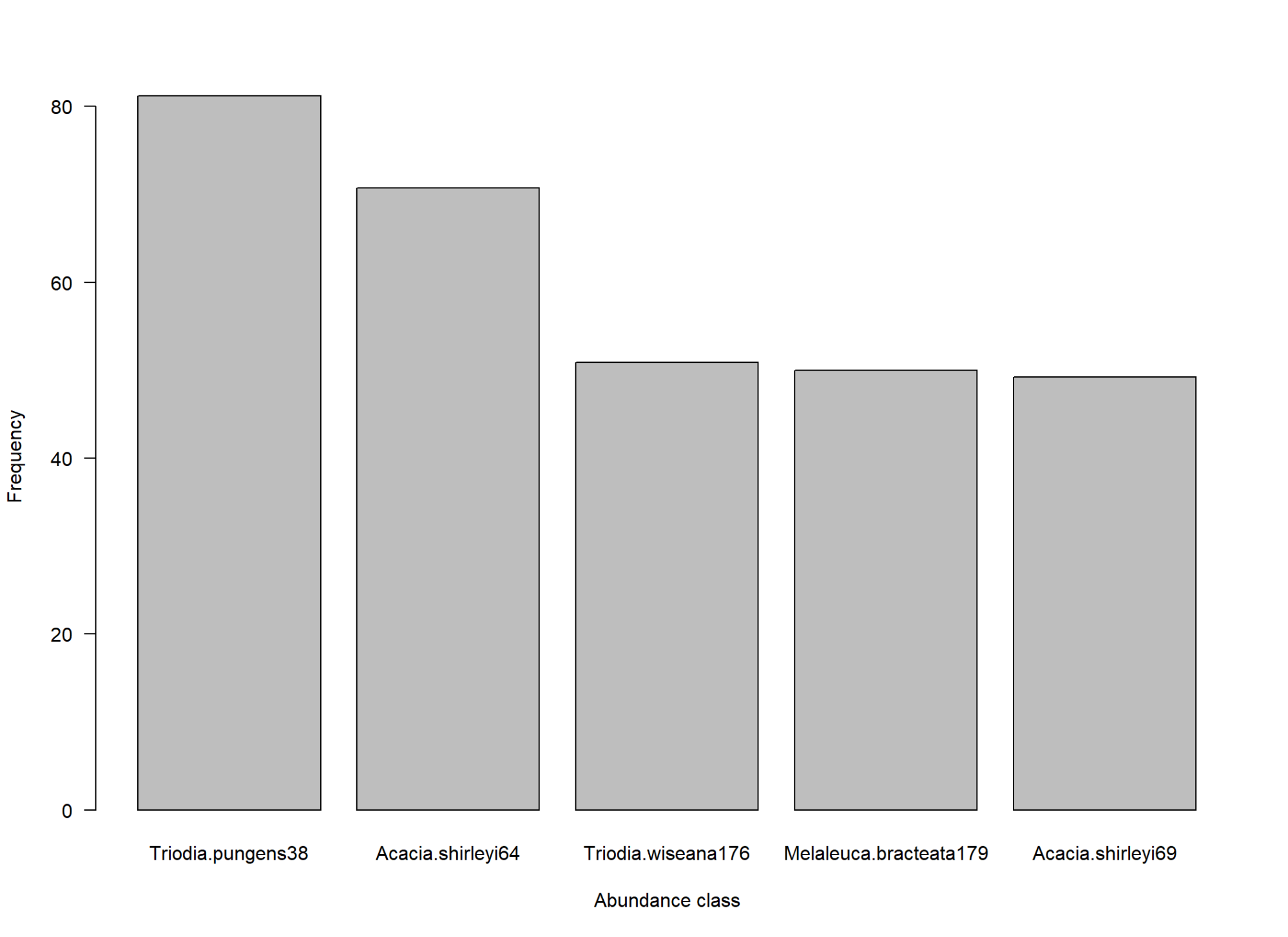
# Minimum and Maximum Site Visit x Species Abundance values  
# ---------------------------------------------------------  
# '-4' because we added 4 new columns (Plot, bioregion.f, longitude, and latitude)##  
range(SppBYSites.BioregTop5[,1: (dim(SppBYSites.BioregTop5)[2]-4)])

## [1] 0.00000 81.18812

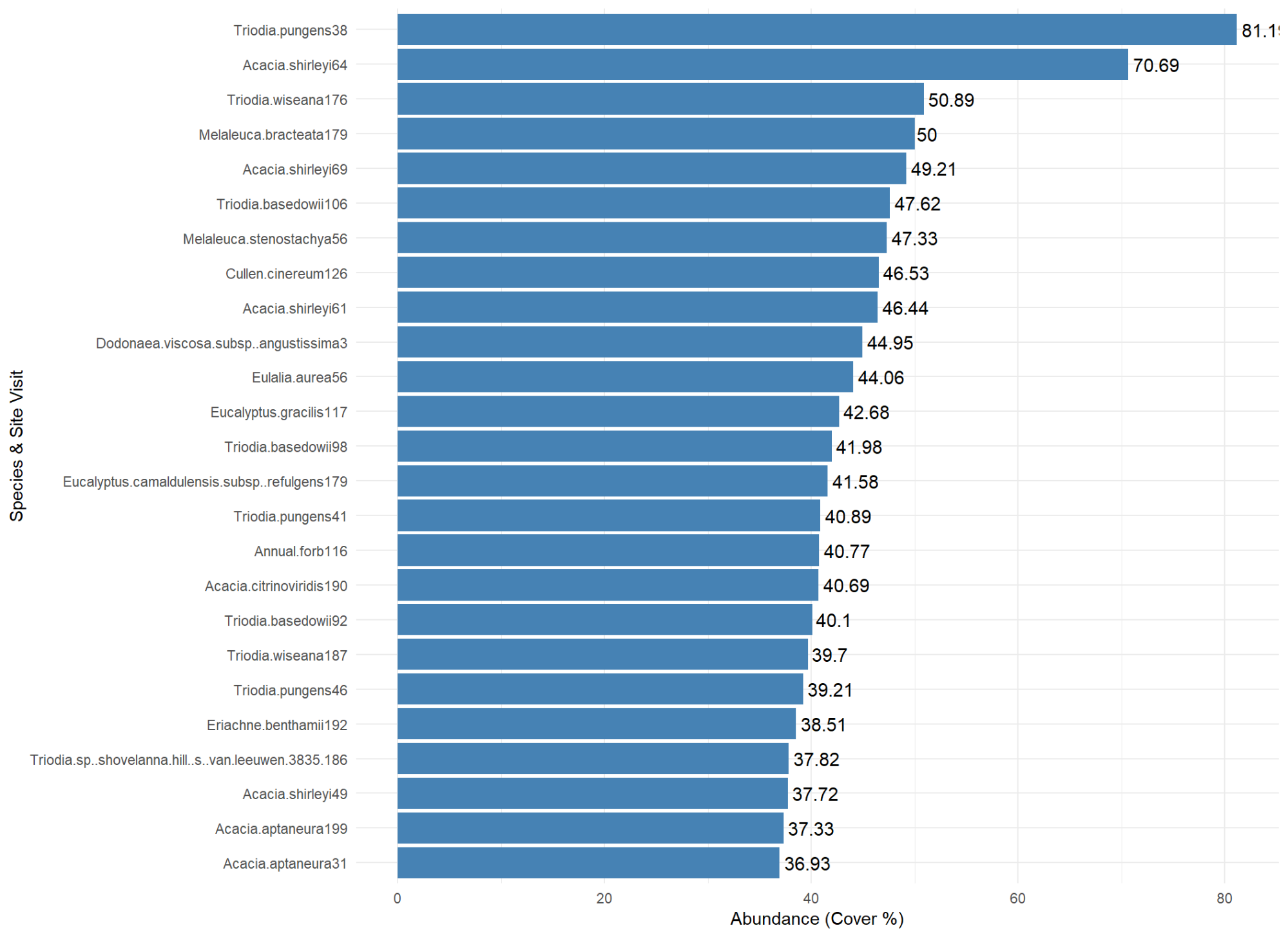
# Plot Highest Site Visit x Species Abundance values  
# ---------------------------------------------------  
Abundance = unlist(SppBYSites.BioregTop5[,1: (dim(SppBYSites.BioregTop5)[2]-4)])  
head(Abundance)

## 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="Frequency")



# ggplot2 graph to make it look nicer. Now we plot the 25 species-site visit covers with horizontal 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 bottom)  
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   
## 0.2970297 1.2871287 2.5757352   
## Abutilon.oxycarpum Abutilon.sp. Acacia.adoxa   
## 0.3960396 0.5944519 0.4950495

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

## [1] 366.8182

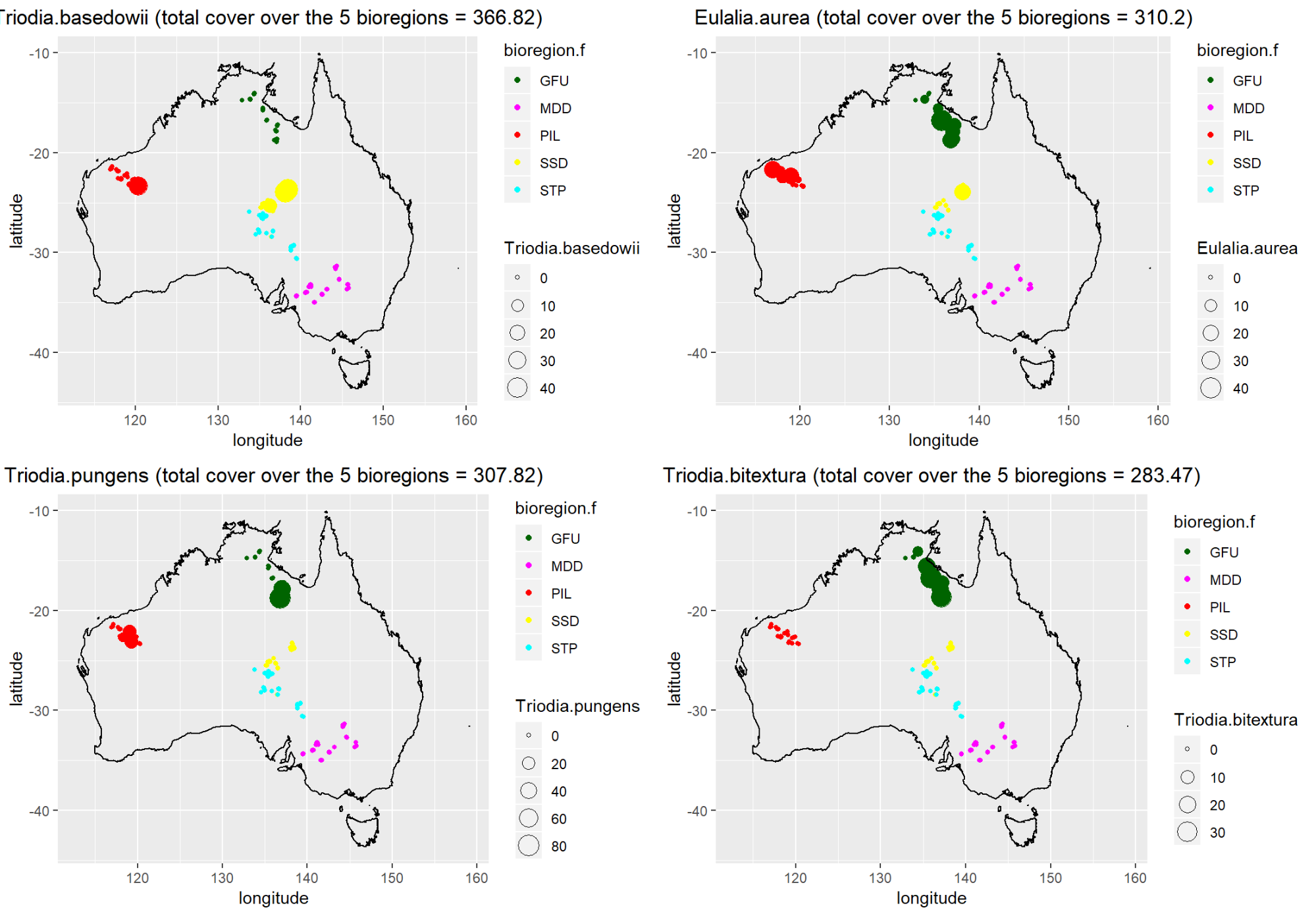
which.max(TotAbundances.BioregTop5)

## Triodia.basedowii   
## 1033

# 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 Eulalia.aurea Triodia.pungens Triodia.bitextura   
## 366.8182 310.1988 307.8218 283.4677

# 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=bioregion.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=bioregion.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=bioregion.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=bioregion.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 Diversity

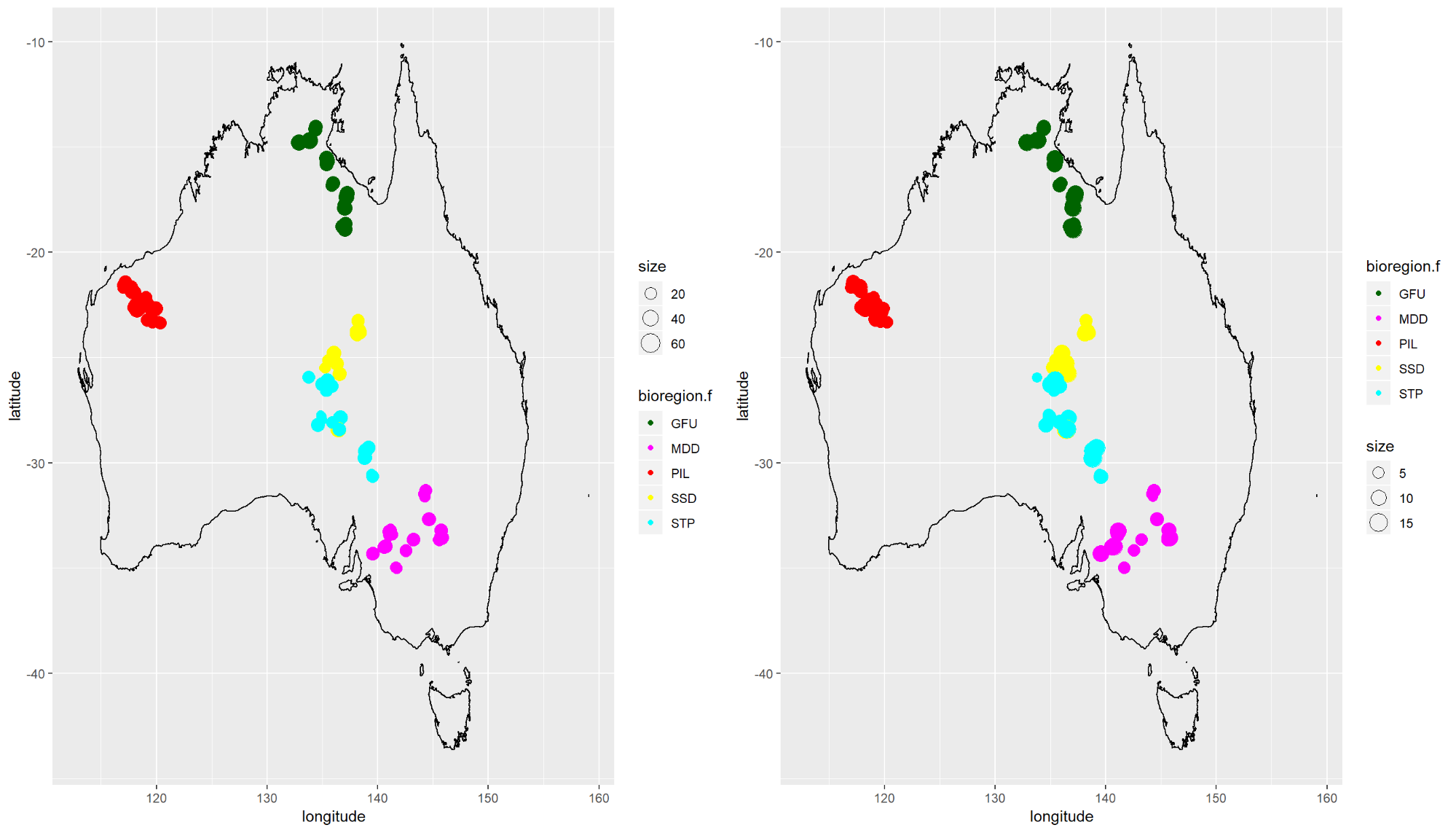
On our exploration of the use of Species-lelvel AusPlots data, we now focus on Species Diversity. We first compute 7 common diversity indices, which we then place in a dataset. Finally, as an example, we plot two of these indices (Species Richness and Shanon Diversity Index) for the sites in the 5 most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

# 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.:1.401 1st Qu.: 4.060 1st Qu.: 2.773   
## Median :18.00 Median :1.779 Median : 5.924 Median : 4.119   
## Mean :19.25 Mean :1.782 Mean : 6.940 Mean : 4.648   
## 3rd Qu.:24.00 3rd Qu.:2.193 3rd Qu.: 8.960 3rd Qu.: 5.778   
## Max. :62.00 Max. :2.866 Max. :17.573 Max. :13.797   
##   
## E1 E2 J bioregion.f  
## Min. :0.08624 Min. :0.05508 Min. :0.1727 GFU:41   
## 1st Qu.:0.26167 1st Qu.:0.16820 1st Qu.:0.5351 MDD:49   
## Median :0.34434 Median :0.23156 Median :0.6363 PIL:35   
## Mean :0.37419 Mean :0.26225 Mean :0.6212 SSD:46   
## 3rd Qu.:0.44732 3rd Qu.:0.30748 3rd Qu.:0.7185 STP:38   
## Max. :1.00000 Max. :1.00000 Max. :0.9610   
## NA's :1   
## longitude latitude   
## Min. :117.0 Min. :-35.00   
## 1st Qu.:134.9 1st Qu.:-29.76   
## Median :136.7 Median :-25.11   
## Mean :134.8 Mean :-25.35   
## 3rd Qu.:139.2 3rd Qu.:-21.87   
## Max. :145.8 Max. :-14.05   
##

# Create Species Richness Plot  
Div.SR =   
ggplot(data=SppBYSites.BioregTop.Div.df, aes(x=longitude, y=latitude, colour=bioregion.f, fill=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, fill=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)



PROPORTIONAL VEGETATION COVER (= FRACTIONAL COVER): `fractional\_cover’ function

Fractional Cover (FC) is the proportional cover of green vegetation, dead vegetation and bare substrate, based on plot-based point intercept data from AusPlots (as generated by ‘get\_ausplots’).

Cover fractions are assigned according to the following:

* ‘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 substrate that is rock, outcrop, gravel or bare soil with no veg cover.

A height rule is applied so that coding to green/brown/bare of the uppermost substrate/vegetation stratum hit at a given point intercept location overrides the others, that is, a dead tree overrides a living shrub beneath and vice versa; substrate coding is overridden by any vegetation cover etc. This means for each of the (usually) 1010 intercepts, there is a single coding and percentage is the number of hits assigned to each fraction, divided by the total number of PIs taken (usually 1010 but can vary) times 100.

There is an option via argument ‘ground\_fractional’ to calculate fractional ground cover - the same concept applied to only 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.

‘In canopy sky’ is excluded by default (only the substrate is considered for those hits) and applies only to regular fractional cover (as trees are excluded in the green fraction for ground fractional cover by default).

Currently, cryptogam substrate is assigned to the non-photosynthetic fraction.

Occasionally substrate type was not collected (‘NC’) or could not be assigned to one of the above categories (‘Unknwn’), in which case a percent cover will be returned under an ‘NA’ fraction if there was no veg cover above those points.

The function fractional\_coverreturns a data frame in which plots are rows, columns are fractions (bare, brown, green and NA) and values are percent cover.

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), ])  
#AP.200Locs.FC = na.omit(AP.200Loc.FC)  
head(AP.200Locs.FC)

## Plot bare brown green NA.  
## NSABBS0002-58555 NSABBS0002-58555 17.13 42.18 40.59 0.1  
## NSABBS0004-58556 NSABBS0004-58556 3.07 34.95 61.98 0.0  
## NSABHC0001-53596 NSABHC0001-53596 31.26 42.24 26.51 0.0  
## NSABHC0002-53597 NSABHC0002-53597 0.00 0.00 30.10 69.9  
## NSABHC0004-53599 NSABHC0004-53599 26.63 36.63 36.73 0.0  
## NSABHC0005-53600 NSABHC0005-53600 20.30 32.77 46.93 0.0

# Create a site\_visit variable in original dataset to relate both datasets  
# ------------------------------------------------------------------------  
AP.data$site.info$Plot = paste(AP.data$site.info$site\_location\_name,   
 AP.data$site.info$site\_location\_visit\_id, sep="-")   
 # Enrich with: Bioregion, Latitude, and Longitude   
AP.200Locs.FC = merge(AP.200Locs.FC, AP.data$site.info, by="Plot")[,c("Plot", "bare", "brown", "green", "NA.", "bioregion.f", "longitude", "latitude")]  
AP.200Locs.FC = na.omit(AP.200Locs.FC)  
head(AP.200Locs.FC)

## Plot bare brown green NA. bioregion.f longitude latitude  
## 1 NSABBS0002-58555 17.13 42.18 40.59 0.1 BBS 148.9925 -31.27945  
## 2 NSABBS0004-58556 3.07 34.95 61.98 0.0 BBS 149.2469 -30.46096  
## 3 NSABHC0001-53596 31.26 42.24 26.51 0.0 BHC 141.4338 -31.89335  
## 4 NSABHC0002-53597 0.00 0.00 30.10 69.9 BHC 141.4330 -31.92703  
## 5 NSABHC0004-53599 26.63 36.63 36.73 0.0 BHC 141.9481 -31.84728  
## 6 NSABHC0005-53600 20.30 32.77 46.93 0.0 BHC 141.9405 -31.84496

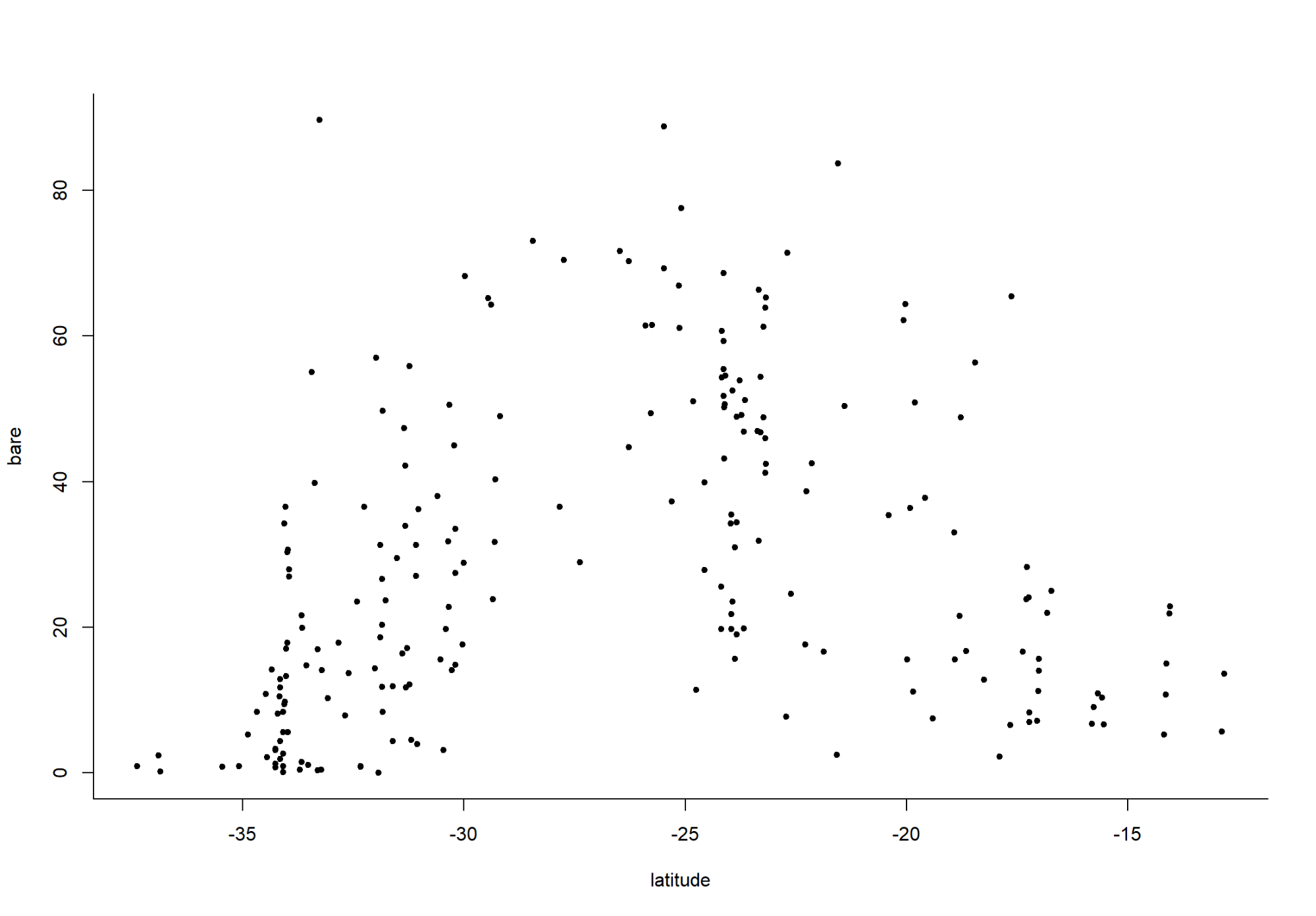
summary(AP.200Locs.FC)

## Plot bare brown green   
## Length:220 Min. : 0.00 Min. : 0.00 Min. : 1.29   
## Class :character 1st Qu.:10.77 1st Qu.:22.00 1st Qu.:21.98   
## Mode :character Median :23.47 Median :30.69 Median :37.43   
## Mean :28.55 Mean :32.08 Mean :38.90   
## 3rd Qu.:46.76 3rd Qu.:41.91 3rd Qu.:51.93   
## Max. :89.70 Max. :77.82 Max. :94.52   
##   
## NA. bioregion.f longitude latitude   
## Min. : 0.0000 SSD : 24 Min. :115.9 Min. :-37.39   
## 1st Qu.: 0.0000 RIV : 22 1st Qu.:133.6 1st Qu.:-32.47   
## Median : 0.0000 GFU : 18 Median :138.3 Median :-27.56   
## Mean : 0.4661 MDD : 17 Mean :136.6 Mean :-26.90   
## 3rd Qu.: 0.0000 BHC : 15 3rd Qu.:141.1 3rd Qu.:-23.18   
## Max. :69.9000 GUP : 13 Max. :151.1 Max. :-12.82   
## (Other):111

names(AP.200Locs.FC)

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

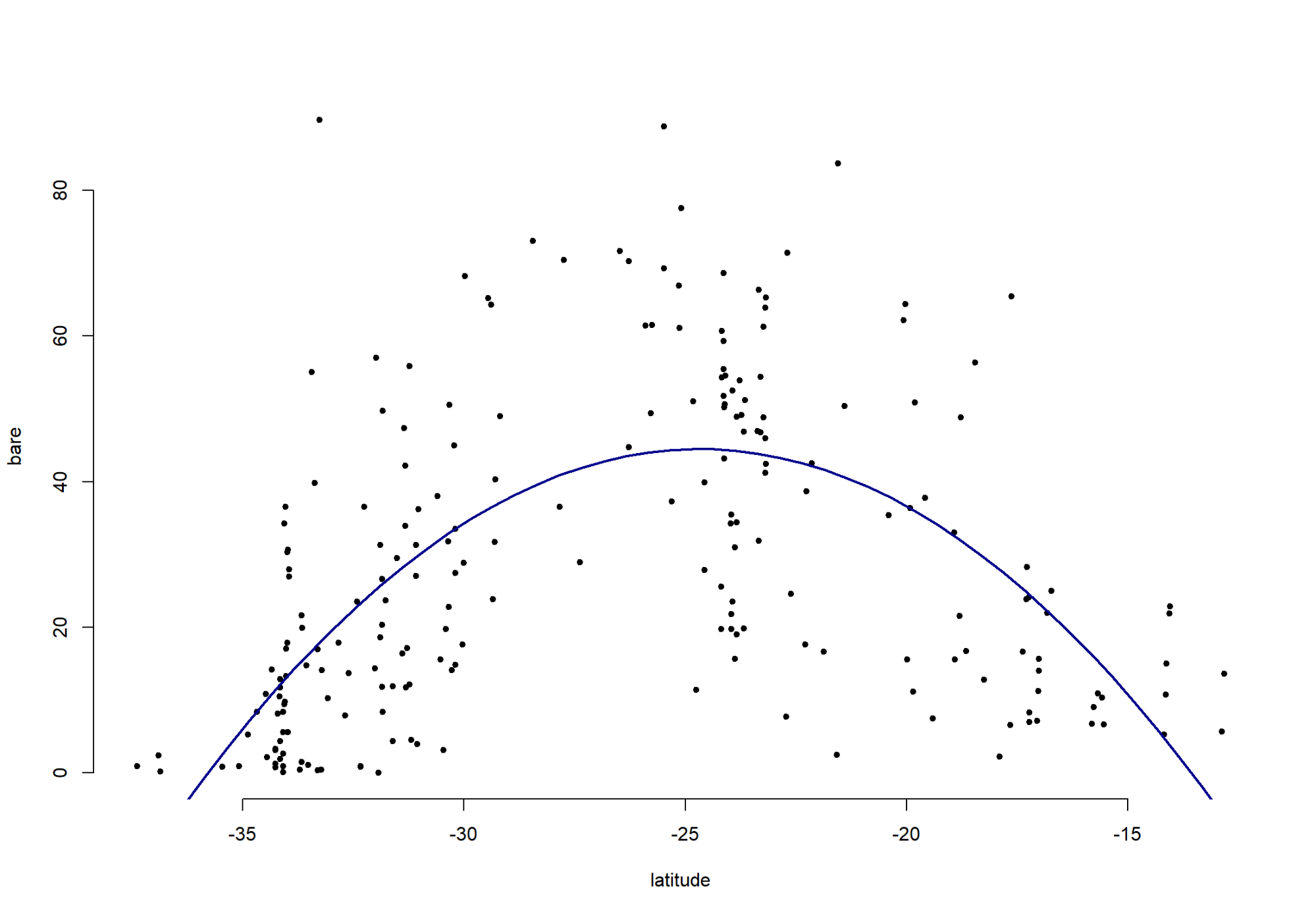
# Plot out the continental relationship between Fractional Cover   
# --------------------------------------------------------------  
# Here we use the 'Proportion of Bare Ground' & Latitude  
  
# Plot the relationship between Proportion of Bare Ground (with no kind of vegetation cover above) and Latitude.  
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 Fractional 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   
## -38.525 -11.903 -1.223 10.571 71.785   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -174.90906 20.72884 -8.438 4.57e-15 \*\*\*  
## latitude -17.78119 1.67620 -10.608 < 2e-16 \*\*\*  
## I(latitude^2) -0.36029 0.03241 -11.117 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 17.16 on 217 degrees of freedom  
## Multiple R-squared: 0.3916, Adjusted R-squared: 0.386   
## F-statistic: 69.83 on 2 and 217 DF, p-value: < 2.2e-16

# Predict values from Model Fit  
pred.df = data.frame(latitude=seq(from=min(AP.200Locs.FC$latitude), to=max(AP.200Locs.FC$latitude), 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 Ground 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")



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 NVIS plant growth forms in plots. The input for this function is a data frame of raw point intercept AusPlots data generated using the get\_ausplots function. Three metrics can be selected to score species growth form:

* *Presence/Absence* (argument m\_kind = PA).
* *Percent Cover*: Based on total frequency of hits (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*: (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 sometimes 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.

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

* *Projected Foliage Cover* (*PFC*):Hits scored as ‘in canopy sky’ are removed (argument cover\_type = PFC).
* *Opaque Canopy Cover* (*OCC*): Hits scored as ‘in canopy sky’ are retained (argument cover\_type = OCC).

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: 574 Sites x 19 Growth Forms

## [1] 209 15

head(AP.BRTop5.GrowthFormBYSites)

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

# Enrich DF  
# =========  
  
# Add Plot (Site-Vist)  
# --------------------  
AP.BRTop5.GrowthFormBYSites$Plot = rownames(AP.BRTop5.GrowthFormBYSites)   
  
# Add: Bioregion, Longitude, Latitude  
# -----------------------------------  
# Both DF have different number of rows (again!)  
dim(AP.BRTop5.GrowthFormBYSites)

## [1] 209 16

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

## [1] 213 43

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

## Chenopod Epiphyte Fern Forb   
## Min. : 0.00 Min. :0.0000 Min. :0.00000 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.: 1.480   
## Median : 1.25 Median :0.0000 Median :0.00000 Median : 6.195   
## Mean : 14.04 Mean :0.0518 Mean :0.08369 Mean :14.525   
## 3rd Qu.: 20.99 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:19.333   
## Max. :100.00 Max. :2.9748 Max. :3.22581 Max. :94.268   
## Fungus Hummock.grass NC Rush   
## Min. :0.00000 Min. : 0.00 Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.00000 1st Qu.: 0.00 1st Qu.:0.00000 1st Qu.:0.000000   
## Median :0.00000 Median : 0.00 Median :0.00000 Median :0.000000   
## Mean :0.00201 Mean : 16.24 Mean :0.06417 Mean :0.004679   
## 3rd Qu.:0.00000 3rd Qu.: 22.42 3rd Qu.:0.00000 3rd Qu.:0.000000   
## Max. :0.42017 Max. :183.86 Max. :2.80374 Max. :0.764331   
## Sedge Shrub Shrub.Mallee Tree.Mallee   
## Min. : 0.0000 Min. : 0.000 Min. : 0.0000 Min. : 0.000   
## 1st Qu.: 0.0000 1st Qu.: 2.488 1st Qu.: 0.0000 1st Qu.: 0.000   
## Median : 0.0000 Median : 9.091 Median : 0.0000 Median : 0.000   
## Mean : 0.7157 Mean :19.164 Mean : 0.6826 Mean : 6.805   
## 3rd Qu.: 0.0000 3rd Qu.:28.101 3rd Qu.: 0.0000 3rd Qu.: 0.000   
## Max. :21.6561 Max. :91.042 Max. :37.7863 Max. :91.979   
## Tree.Palm Tussock.grass Vine Plot   
## Min. : 0.0000 Min. : 0.000 Min. : 0.0000 Length:209   
## 1st Qu.: 0.0000 1st Qu.: 2.262 1st Qu.: 0.0000 Class :character   
## Median : 0.3745 Median :16.425 Median : 0.0000 Mode :character   
## Mean :12.3747 Mean :27.278 Mean : 0.3843   
## 3rd Qu.:15.1351 3rd Qu.:45.905 3rd Qu.: 0.0000   
## Max. :99.1747 Max. :99.051 Max. :26.1456   
## bioregion.f longitude latitude   
## GFU:41 Min. :117.0 Min. :-35.00   
## MDD:49 1st Qu.:134.9 1st Qu.:-29.76   
## PIL:35 Median :136.7 Median :-25.11   
## SSD:46 Mean :134.8 Mean :-25.35   
## STP:38 3rd Qu.:139.2 3rd Qu.:-21.87   
## Max. :145.8 Max. :-14.05

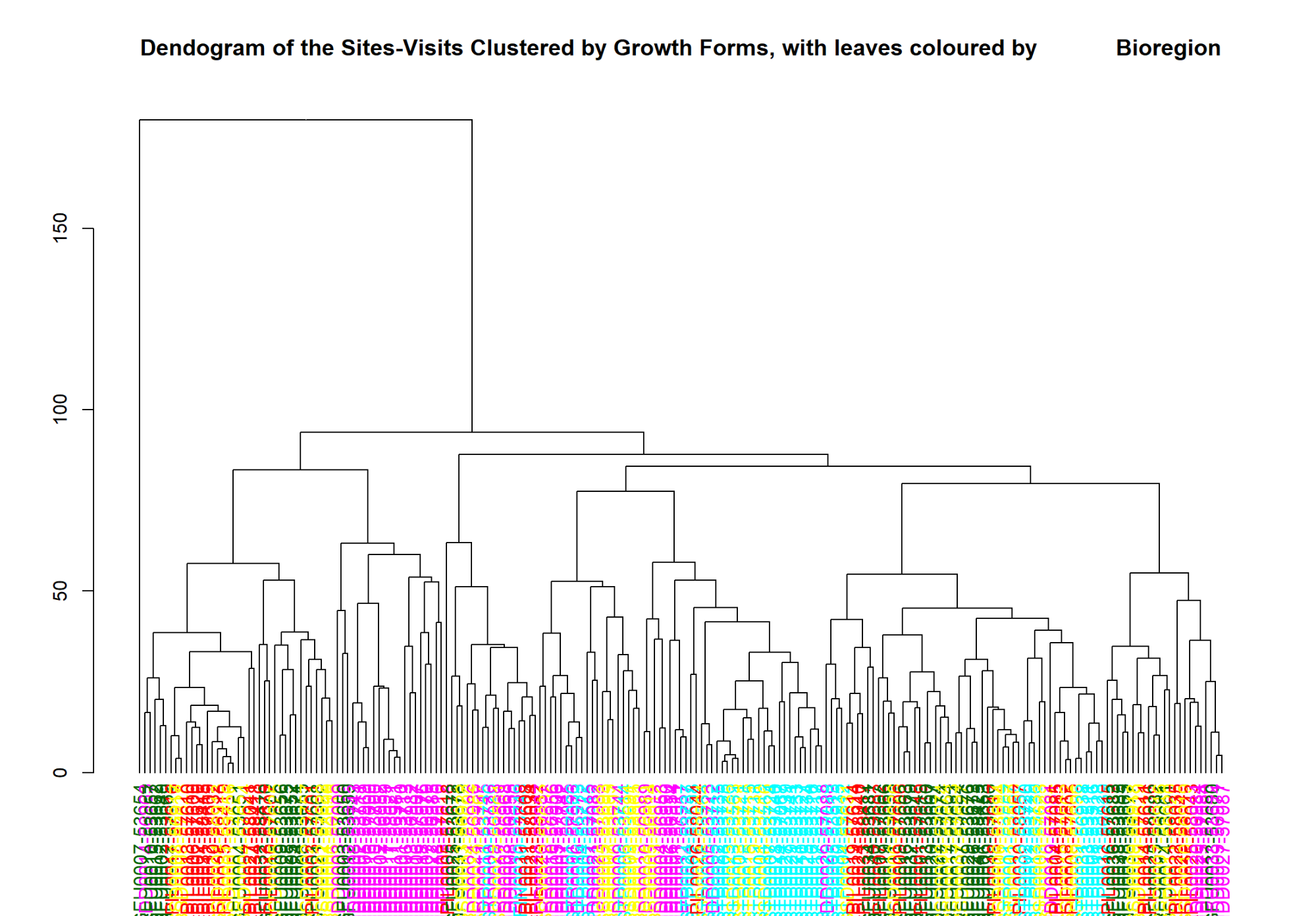
names(AP.BRTop5.GrowthFormBYSites)

## [1] "Chenopod" "Epiphyte" "Fern" "Forb"   
## [5] "Fungus" "Hummock.grass" "NC" "Rush"   
## [9] "Sedge" "Shrub" "Shrub.Mallee" "Tree.Mallee"   
## [13] "Tree.Palm" "Tussock.grass" "Vine" "Plot"   
## [17] "bioregion.f" "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  
## Chenopod 0 1.25 100.00 14.04 22.35  
## Epiphyte 0 0.00 2.97 0.05 0.28  
## Fern 0 0.00 3.23 0.08 0.39  
## Forb 0 6.19 94.27 14.53 18.96  
## Fungus 0 0.00 0.42 0.00 0.03  
## Hummock.grass 0 0.00 183.86 16.24 29.09  
## NC 0 0.00 2.80 0.06 0.33  
## Rush 0 0.00 0.76 0.00 0.05  
## Sedge 0 0.00 21.66 0.72 2.83  
## Shrub 0 9.09 91.04 19.16 22.43  
## Shrub.Mallee 0 0.00 37.79 0.68 4.04  
## Tree.Mallee 0 0.00 91.98 6.81 17.62  
## Tree.Palm 0 0.37 99.17 12.37 21.88  
## Tussock.grass 0 16.43 99.05 27.28 27.82  
## Vine 0 0.00 26.15 0.38 2.06

# 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$Plot  
  
# 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 codes bellow should be revised.  
# Here the codes correspond to: MDD (Murry Darling Depression), SSD (Simpson   
# Strzelecki Dunefields), GFU (Gulf Fall and Uplands), STP (Stony Plains),   
# PIL (Pilbara)  
AP.BRTop5.GrowthFormBYSites$bioregion.col.f = AP.BRTop5.GrowthFormBYSites$bioregion.f  
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f) == "GFU"] = "darkgreen"  
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f) == "MDD"] = "magenta"  
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f) == "PIL"] = "red"  
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f) == "SSD"] = "yellow"  
levels(AP.BRTop5.GrowthFormBYSites$bioregion.col.f)[levels(AP.BRTop5.GrowthFormBYSites$bioregion.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  
plot(AP.BRTop5.GFBYSites.dend,   
 main="Dendogram of the Sites-Visits Clustered by Growth Forms, with leaves coloured by Bioregion")



BASAL AREA (OR NUMBER OF BASAL WEDGE HITS): basal\_area function (for 5 most sampled bioregions).

The basal\_area function calculates the Basal Area (or Number of Basal Wedge Hits) for each plot, using the raw basal wedge data returned by the get\_ausplots funcion also in the ausplotsR package. This function returns a data frame with rows representing Plots (or species by plots) and a single column containning the Basal Area (m2/ha) or Hit Scores.

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:99 Min. : 0.2857   
## Class :character 1st Qu.: 1.4333   
## Mode :character Median : 3.6611   
## Mean : 4.4389   
## 3rd Qu.: 5.3542   
## Max. :15.8000

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"

colnames(AP.BioregTop5.BA)[1] = "Plot" # Change the name to be able to Merge DFs  
summary(AP.BioregTop5.BA)

## Plot basal\_area\_m2\_ha   
## Length:99 Min. : 0.2857   
## Class :character 1st Qu.: 1.4333   
## Mode :character Median : 3.6611   
## Mean : 4.4389   
## 3rd Qu.: 5.3542   
## Max. :15.8000

head(AP.BioregTop5.BA)

## Plot 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 (again!)  
dim(AP.BioregTop5.BA)

## [1] 99 2

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

## [1] 213 43

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

## Plot 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

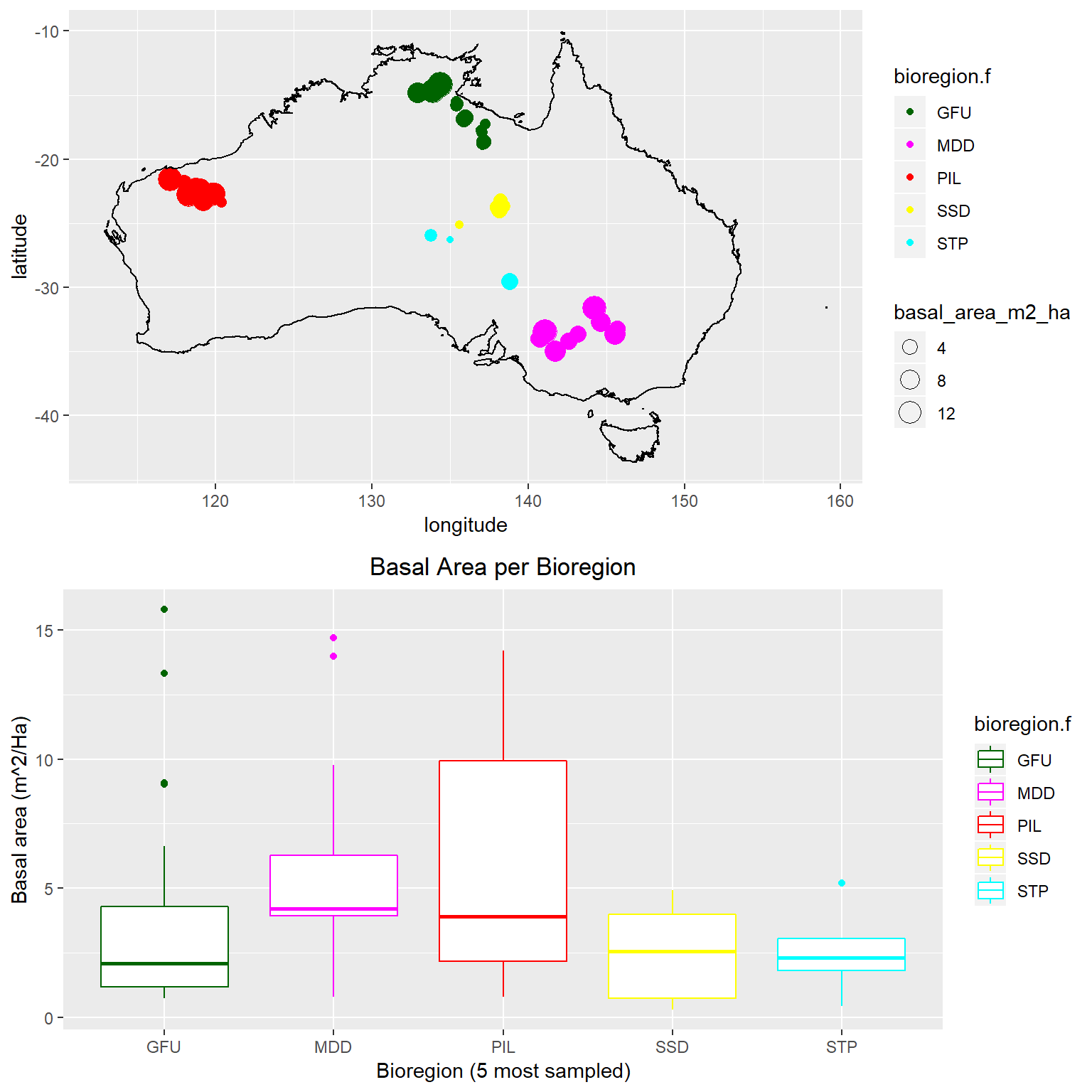
summary(AP.BioregTop5.BA)

## Plot basal\_area\_m2\_ha bioregion.f longitude   
## Length:99 Min. : 0.2857 GFU:34 Min. :117.1   
## Class :character 1st Qu.: 1.4333 MDD:27 1st Qu.:126.6   
## Mode :character Median : 3.6611 PIL:25 Median :136.0   
## Mean : 4.4389 SSD: 9 Mean :133.5   
## 3rd Qu.: 5.3542 STP: 4 3rd Qu.:140.6   
## Max. :15.8000 Max. :145.8   
## latitude   
## Min. :-35.00   
## 1st Qu.:-31.55   
## Median :-22.62   
## Mean :-23.65   
## 3rd Qu.:-17.82   
## Max. :-14.05

names(AP.BioregTop5.BA)

## [1] "Plot" "basal\_area\_m2\_ha" "bioregion.f"   
## [4] "longitude" "latitude"

# Graphical Visualisation  
# =======================  
  
# Map with circle size the Basal Area (m2/ha)  
# -------------------------------------------  
AP.BioregTop5.BA.p1 =   
ggplot(data=AP.BioregTop5.BA, aes(x=longitude, y=latitude, colour=bioregion.f, fill=bioregion.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", "yellow", "cyan")) +  
labs(title="Basal Area per Bioregion",x="Bioregion (5 most sampled)", y = "Basal area (m^2/Ha)") +   
theme(plot.title = element\_text(hjust = 0.5))  
  
# Plot both graphs  
# ----------------  
grid.arrange(AP.BioregTop5.BA.p1, AP.BioregTop5.BA.p2, nrow=2)



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