

# colrow: Handling SimU, CR, and LU data

*Pedro R. Andrade*

*25 September 2018*

## Abstract

This package has functions to handle Simulation Units (SimU), ColRow (CR), and Large Unit (LU) data. It shows examples on how to export them as shapefiles, create maps, as well as summarize and plot them over time.

## Introduction

This package allows users to handle CR, SimU, and LU data. In order to start, it is necessary to install the package from GitHub.

```
devtools::install_github("pedro-andrade-inpe/colrow")
```

If the package will be installed in a Linux machine, it is necessary to manually install some dependencies. Please visit the [webpage of the package](#) in GitHub for more details.

## Creating shapefiles for a given country

First load the package.

```
require(colrow)
```

```
## Loading required package: colrow
```

```
## colrow - Handling SimU, CR, and LU data
```

```
## Version 0.2-0 is now loaded
```

Download world data from Dropbox [here](#) and save them in a single directory. Create a variable in R to store this directory.

```
dataDir <- "c:/Users/pedro/Dropbox/colrow/"
```

The content of the directory should have 17 files as shown below.

```
list.files(dataDir)
```

```
## [1] "COLROW30.aux"      "COLROW30.dbf"      "COLROW30.sbn"
## [4] "COLROW30.sbx"      "COLROW30.shp"      "COLROW30.shx"
## [7] "g2006_2.dbf"       "g2006_2.shp"       "g2006_2.shp.xml"
## [10] "g2006_2.shx"       "SimU_all.dbf"      "SimU_all.prj"
## [13] "SimU_all.sbn"      "SimU_all.sbx"      "SimU_all.shp"
## [16] "SimU_all.shp.xml"  "SimU_all.shx"
```

See the available countries in alphabetical order, use `getCountries()`. The code below shows the first 10.

```
colrow::getCountries(dataDir)[1:10]
```

```
## [1] "Afghanistan"      "Aksai Chin"        "Albania"
## [4] "Algeria"           "American Samoa"    "Andorra"
## [7] "Angola"            "Anguilla"          "Antarctica"
## [10] "Antigua and Barbuda"
```

Then use `getLU()`, `getCR()`, and/or `getSimU()` to generate geospatial data for a given country. They get as first argument the country name written in the same way as one of the outputs of `getCountries()`. The second argument is the directory where the data downloaded from Dropbox was stored. These functions take some time to execute.

```
country <- "Brazil"
```

```
myLU <- colrow::getLU(country, dataDir)
```

```
## Reading all countries
## Selecting Brazil
## Reading all SimUs
## Subsetting SimUs
## Mapping SimU to LU
## Merging data
## Computing union of SimUs within the same LU
```

```
myCR <- colrow::getCR(country, dataDir)
```

```
## Reading all countries
## Selecting Brazil
## Reading all SimUs
## Subsetting SimUs
## Mapping SimU to CR
```

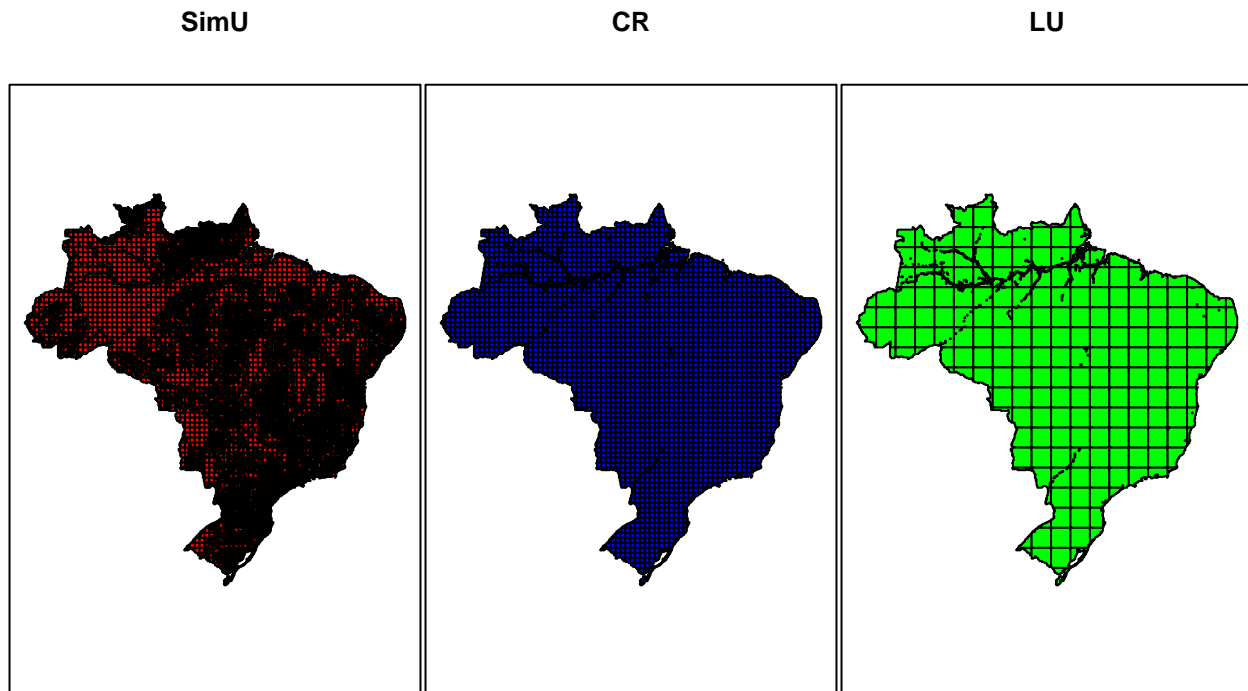
```
mySimU <- colrow::getSimU(country, dataDir)
```

```
## Reading all countries
## Selecting Brazil
## Reading all SimUs
## Subsetting SimUs
## Joining SimUs
```

To see that everything is correct, plot the data.

```
par(mfrow = c(1, 3), mar = c(5, 0.1, 5, 0.1))
```

```
plot(sf::st_geometry(mySimU), main = "SimU", col = "red"); box()
plot(sf::st_geometry(myCR),   main = "CR", col = "blue"); box()
plot(sf::st_geometry(myLU),   main = "LU", col = "green"); box()
```



Finally, save the data to be used by the processing functions. Each of this files has a column named ID that will be used to match the objects from the CSV data. Once these files are saved, it is not necessary to create them again every time outputs are processed.

```
sf::write_sf(myLU, paste0(country, "LU.shp"))
sf::write_sf(myCR, paste0(country, "CR.shp"))
sf::write_sf(mySimU, paste0(country, "SimU.shp"))
```

## Processing outputs

To process a CSV file, it is necessary to choose the created shapefile that matches the representation used in the csv. Then, it is necessary to define the attribute names, as the CSV files do not store a header. These names must have one attribute called ID and another called VALUE. Function `processFile()` gets the shapefile, the CSV file, and the description of the attributes as arguments. You can use `colrow::attrs()` to avoid quotes.

```
csvfile <- system.file("extdata/scenarios/FC/Land_Compare3_FC.csv", package = "colrow")
attributes <- colrow::attrs(COUNTRY, ID, ALTI, SLP, SOIL, USE, SCENARIO, YEAR, VALUE)

result <- colrow::processFile(
  "BrazilCR.shp",
  csvfile,
  attributes
)
```

```
## Reading shapefile: BrazilCR.shp
```

```
## Reading data file: C:/Users/pedro/Documents/R/win-library/3.5/colrow/extdata/scenarios/FC/Land_Compa
```

```
## 6 objects belong to the shapefile but not to the csv file: CR273185, CR273186, CR274185, CR274186, C
## Ignored attributes: COUNTRY, ALTI, SLP, SOIL, SCENARIO
## Attributes to be joined: USE, YEAR
## USE: CrpLnd, PriFor, NatLnd, ForReg, GrsLnd, MngFor, PltFor
## YEAR: 2000, 2010, 2020, 2030, 2040, 2050
## Spreading the data
## 40 attributes were created
## Merging the data
## Replacing 34340 NA values by zero
```

This function automatically ignores all attributes that have only one value (COUNTRY, ALTI, SLP, SOIL and SCENARIO above). The other attributes (except ID and VALUE) compose the created attribute names (USE and YEAR above). They are concatenated according to their order in the CSV file.

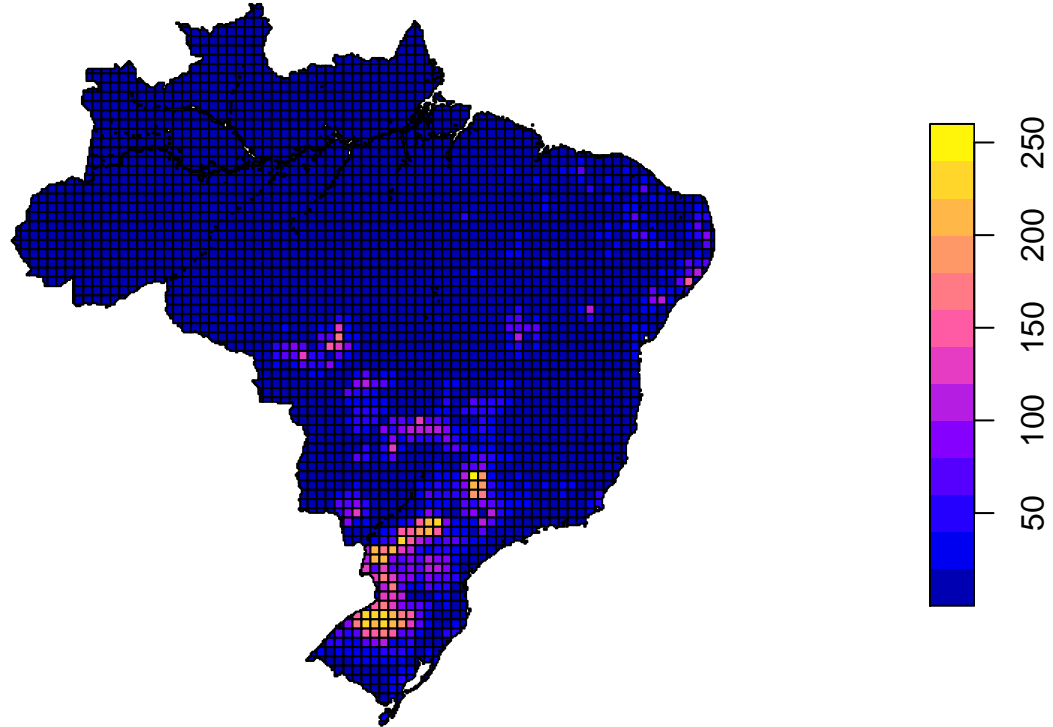
```
names(result)
```

```
## [1] "ID" "CrpLnd2000" "CrpLnd2010" "CrpLnd2020" "CrpLnd2030"
## [6] "CrpLnd2040" "CrpLnd2050" "ForReg2030" "ForReg2040" "ForReg2050"
## [11] "GrsLnd2000" "GrsLnd2010" "GrsLnd2020" "GrsLnd2030" "GrsLnd2040"
## [16] "GrsLnd2050" "MngFor2000" "MngFor2010" "MngFor2020" "MngFor2030"
## [21] "MngFor2040" "MngFor2050" "NatLnd2000" "NatLnd2010" "NatLnd2020"
## [26] "NatLnd2030" "NatLnd2040" "NatLnd2050" "PltFor2000" "PltFor2010"
## [31] "PltFor2020" "PltFor2030" "PltFor2040" "PltFor2050" "PriFor2000"
## [36] "PriFor2010" "PriFor2020" "PriFor2030" "PriFor2040" "PriFor2050"
## [41] "geometry"
```

It is possible to plot the created attributes by simply calling `plot()`, selecting a given attribute.

```
plot(result["CrpLnd2000"])
```

## CrpLnd2000



It is also possible to use `processFile()` to save the output when processing a file by adding a fourth argument with the output file name. In this case, this function does not return anything.

```
colrow::processFile(  
  "BrazilCR.shp",  
  csvfile,  
  attributes,  
  "brazillandCompare.shp"  
)
```

```
## Reading shapefile: BrazilCR.shp  
## Reading data file: C:/Users/pedro/Documents/R/win-library/3.5/colrow/extdata/scenarios/FC/Land_Compar  
## 6 objects belong to the shapefile but not to the csv file: CR273185, CR273186, CR274185, CR274186, C  
## Ignored attributes: COUNTRY, ALTI, SLP, SOIL, SCENARIO  
## Attributes to be joined: USE, YEAR  
## USE: CrpLnd, PriFor, NatLnd, ForReg, GrsLnd, MngFor, PltFor  
## YEAR: 2000, 2010, 2020, 2030, 2040, 2050  
## Spreading the data  
## 40 attributes were created  
## Merging the data  
## Replacing 34340 NA values by zero  
## Renaming attributes according to convertList  
## No attribute has more than 10 characters  
## Writing output file: brazillandCompare.shp
```

Note that shapefiles have limits: 255 or less columns, ten or less characters in the attribute names. When

working with large attribute names, it is necessary to simplify the names. In the case below, seven names will be replaced. For instance, `CrpLnd` will be replaced by `cr` in the attribute names.

```
convert <- list(
  CrpLnd = "cr", PriFor = "pr",
  NatLnd = "nl", ForReg = "fr",
  GrsLnd = "gl", MngFor = "mf",
  PltFor = "pl"
)
```

If you want to simplify the years to two characters, just run:

```
for(year in paste(seq(2000, 2050, 10))) # from 2000, 2010, ..., 2050
  convert[[year]] = substr(year, 3, 4) # to 00, 10, ..., 50

unlist(convert)
```

```
## CrpLnd PriFor NatLnd ForReg GrsLnd MngFor PltFor 2000 2010 2020
## "cr" "pr" "nl" "fr" "gl" "mf" "pl" "00" "10" "20"
## 2030 2040 2050
## "30" "40" "50"
```

Finally, these simplifications can be used as fifth argument to `processFile()`.

```
colrow::processFile(
  "BrazilCR.shp",
  csvfile,
  attributes,
  "brazilOutput.shp",
  convert
)
```

```
## Reading shapefile: BrazilCR.shp
## Reading data file: C:/Users/pedro/Documents/R/win-library/3.5/colrow/extdata/scenarios/FC/Land_Compar
## 6 objects belong to the shapefile but not to the csv file: CR273185, CR273186, CR274185, CR274186, CR
## Ignored attributes: COUNTRY, ALTI, SLP, SOIL, SCENARIO
## Attributes to be joined: USE, YEAR
## USE: CrpLnd, PriFor, NatLnd, ForReg, GrsLnd, MngFor, PltFor
## YEAR: 2000, 2010, 2020, 2030, 2040, 2050
## Spreading the data
## 40 attributes were created
## Merging the data
## Replacing 34340 NA values by zero
## Renaming attributes according to convertList
## No attribute has more than 10 characters
## Writing output file: brazilOutput.shp
```

Sometimes there are more than one `VALUE` per unique `ID`, for each unique combination of attributes. In this case, it is necessary to transform all the values into one single result. As default, `processFile()` sums all the values, but it is possible to use any other function passed as argument, such as `mean()`. Note that `sum()` is used for absolute values and `mean()` for averages. To use another function, use argument `aggregate`, passing a function as value, as shown in the example below.

```
result <- colrow::processFile(
  "BrazilCR.shp",
  system.file("extdata/csv/YIELD_COMPARE2.CSV", package = "colrow"),
  colrow::attrs(ID, CROP, ScenYear, VALUE),
```

```

    aggregate = sum # default value, could be omitted
  )

## Reading shapefile: BrazilCR.shp
## Reading data file: C:/Users/pedro/Documents/R/win-library/3.5/colrow/extdata/csv/YIELD_COMPARE2.CSV
## 177 objects belong to the shapefile but not to the csv file: CR214192, CR214193, CR218189, CR220180,
## Ignored attributes:
## Attributes to be joined: CROP, ScenYear
## CROP: Corn, Rice
## ScenYear: 2000, 2010
## Spreading the data
## 5 attributes were created
## Merging the data
## Replacing 1322 NA values by zero

```

## Plotting outputs

The example of plotting in this section uses `tmap` package.

```
require(tmap)
```

```
## Loading required package: tmap
```

In this example, we will draw the Brazilian Biomes on top of the ColRows. The shapefile with this data is stored in file `br_biomes.shp` available within `colrow` package. We will use CR data from `brazilLandCompare.shp`, created previously in this tutorial.

```

brazil <- sf::read_sf("brazilLandCompare.shp")
biomes <- system.file("extdata/shape", "br_biomes.shp", package = "colrow") %>% sf::read_sf()

```

Then select a palette from the functions availables in `RColorBrewer` using `brewer.pal()` (see the description of all colors available [here](#)) and choose the cuts to represent the intervals of each slice. In the code below, as we have seven cuts, we need to define six colors.

```
max(brazil$GrsLnd2010)
```

```
## [1] 303.33
```

```

cuts <- c(0, 50, 100, 150, 200, 250, 305)
rdPu <- RColorBrewer::brewer.pal(6, "RdPu")

```

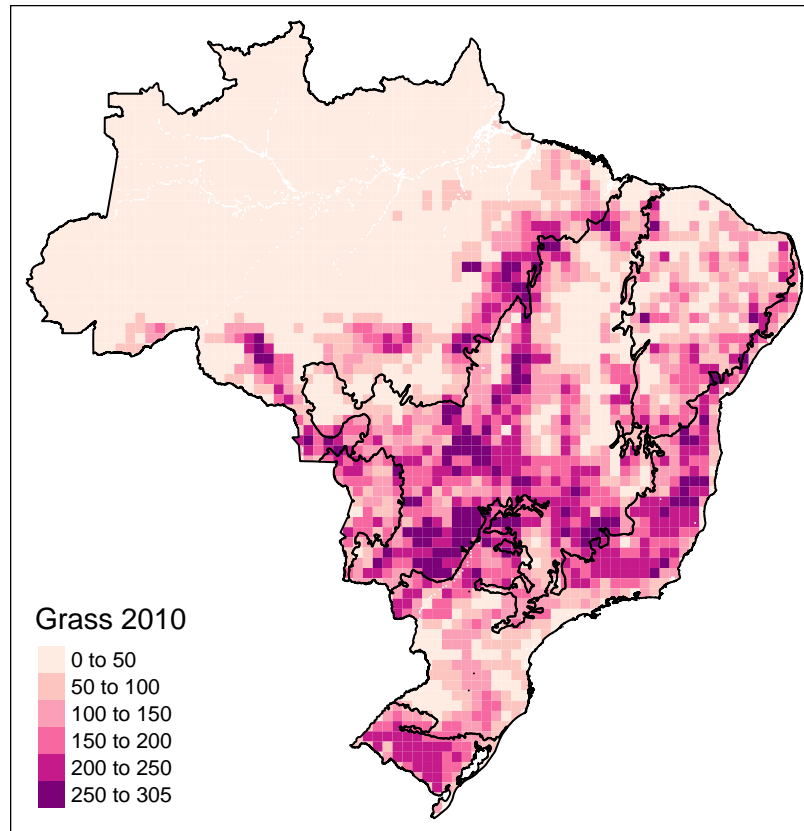
Finally, use `tm_shape()`, `tm_fill()`, and `tm_borders()` to draw the map.

```

tm_shape(brazil) +
  tm_fill(col = "GrsLnd2010", palette = rdPu, breaks = cuts, title = "Grass 2010") +
  tm_shape(biomes) +
  tm_borders(lwd = 1, col = "black")

```

```
## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3
```

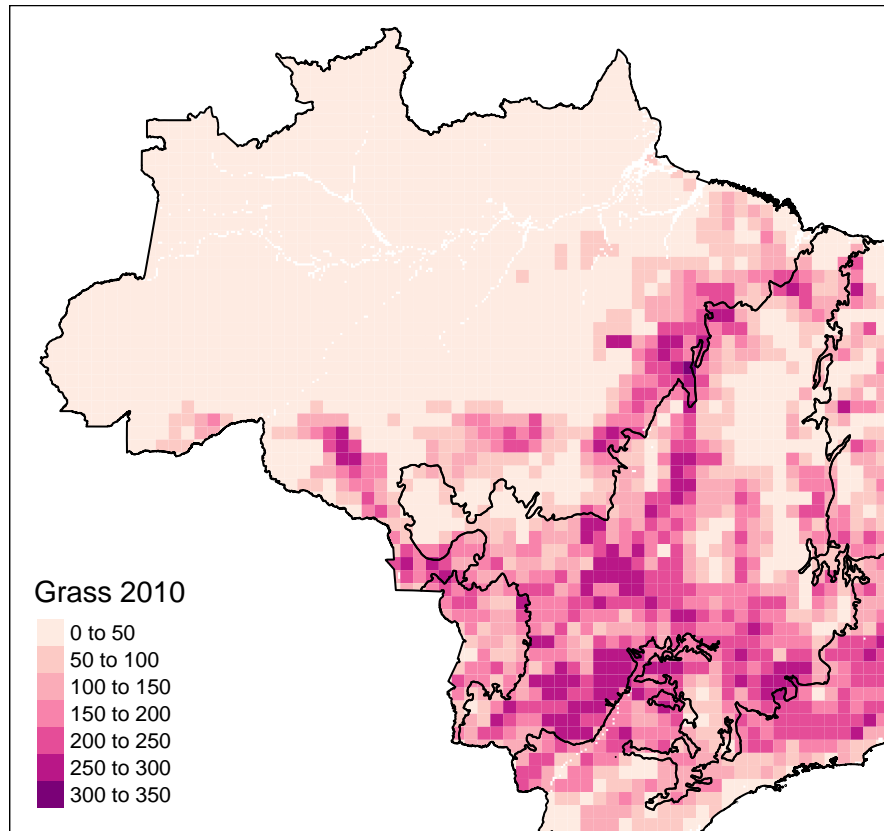


If you want to zoom in the map to a given region, it is possible to set a `bbox` to the `tm_shape()`:

```
amazCerradoBox <- c(xmin = -74.5, xmax = -41.4, ymin = -25, ymax = 5.5) %>%
  sf::st_bbox(crs = st_crs(4326))

tm_shape(brazil, bbox = amazCerradoBox) +
  tm_fill(col = "GrsLnd2010", palette = rdPu, title = "Grass 2010") +
  tm_shape(biomes) +
  tm_borders(lwd = 1, col = "black")
```





## Handling csv files

You can handle csv files directly to create charts or to compute values without creating shapefiles. First read data as a tibble using `readCSV()`:

```
data <- colrow::readCSV(
  system.file("extdata/scenarios/FC/Land_Compare3_FC.csv", package = "colrow"),
  colrow::attrs(COUNTRY, ID, ALTICLASS, SLPCLASS, SOILCLASS, USE, SCENARIO, YEAR, VALUE)
)
```

To get the unique values per class, just use `unique()`:

```
unique(data$USE)
```

```
## [1] "CrpLnd" "PriFor" "NatLnd" "ForReg" "GrsLnd" "MngFor" "PltFor"
```

```
unique(data$YEAR)
```

```
## [1] 2000 2010 2020 2030 2040 2050
```

To check if the sum of all land use areas are the same for each year use some `dplyr` functions:

```
result <- data %>%
  dplyr::group_by(YEAR) %>%
  dplyr::summarise(VALUE = sum(VALUE))
```

```
result
```

```
## # A tibble: 6 x 2
```

```
##   YEAR   VALUE
##   <int>  <dbl>
## 1  2000 584619.
## 2  2010 584619.
## 3  2020 584619.
## 4  2030 584618.
## 5  2040 584618.
## 6  2050 584618.
```

Then let us now compute the total of each land use area per year and rename the attributes in order to use them directly in the plot.

```
result <- data %>%
  dplyr::group_by(YEAR, USE) %>%
  dplyr::summarise(VALUE = sum(VALUE)) %>% # sum by year/use
  dplyr::mutate(Year = YEAR, Use = USE) %>% # rename variables
  dplyr::mutate(Use = dplyr::recode(Use,      # rename attributes
    CrpLnd = "Crop Land",
    ForReg = "Forest Regrowth",
    GrsLnd = "Grass Land",
    MngFor = "Managed Forest",
    NatLnd = "Natural Land",
    PltFor = "Planted Forest",
    PriFor = "Primary Forest"))

result
```

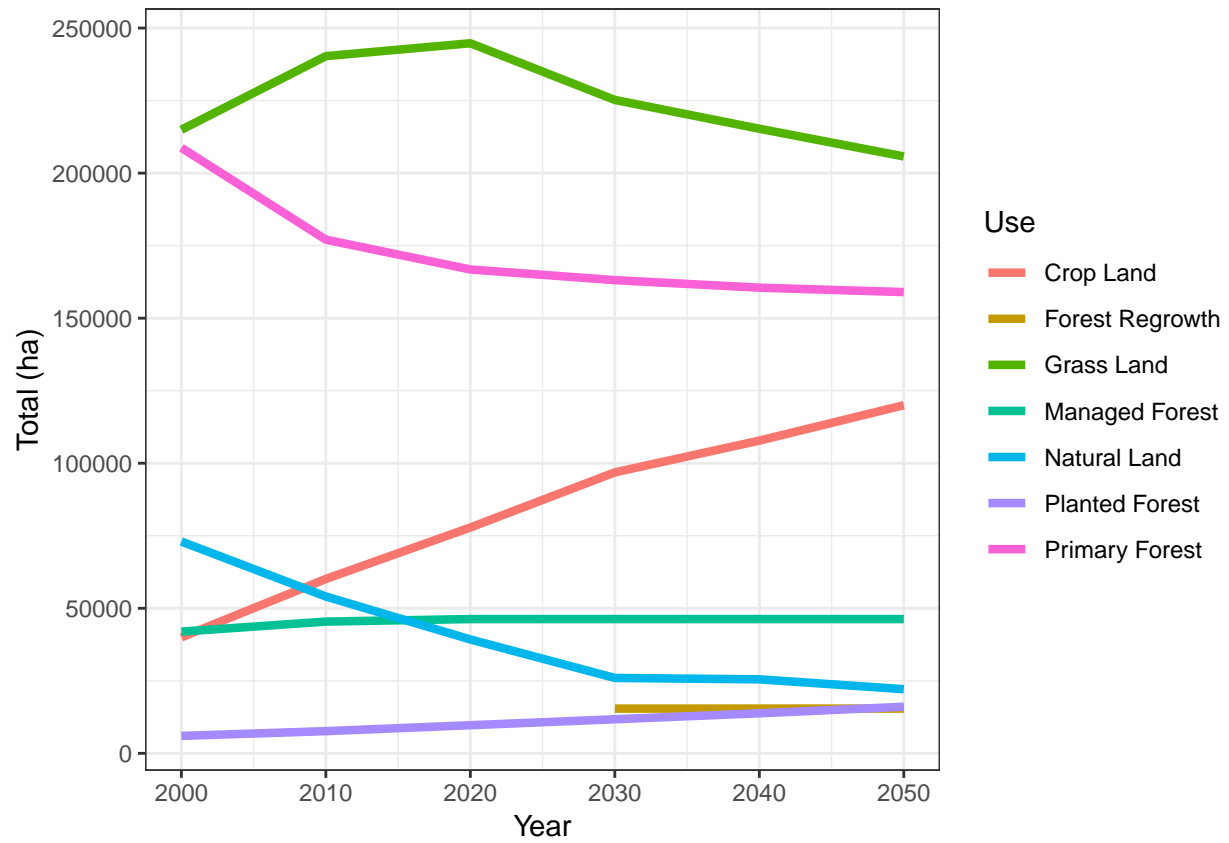
```
## # A tibble: 39 x 5
## # Groups:   YEAR [6]
##   YEAR USE      VALUE Year Use
##   <int> <chr>    <dbl> <int> <chr>
## 1  2000 CrpLnd  39983.  2000 Crop Land
## 2  2000 GrsLnd 214941.  2000 Grass Land
## 3  2000 MngFor  41955.  2000 Managed Forest
## 4  2000 NatLnd  73001.  2000 Natural Land
## 5  2000 PltFor   6005.  2000 Planted Forest
## 6  2000 PriFor 208736.  2000 Primary Forest
## 7  2010 CrpLnd  60109.  2010 Crop Land
## 8  2010 GrsLnd 240326.  2010 Grass Land
## 9  2010 MngFor  45416.  2010 Managed Forest
## 10 2010 NatLnd  54057.  2010 Natural Land
## # ... with 29 more rows
```

Finally, just use ggplot to draw the evolution of each land use area per year.

```
require(ggplot2)

## Loading required package: ggplot2

ggplot(result) +
  aes(x = Year, y = VALUE, colour = Use) +
  geom_line(lwd = 1.5) +
  theme_bw() +
  ylab("Total (ha)")
```



## Final remarks

If you have suggestions or want to report an error, please visit the GitHub page [here](#).