Changes in Land Use between 1990 and 2014

https://github.com/RsMarx/Ag_Forest_Data_Final-

Rebecca Marx

Abstract

Experimental overview. This section should be no longer than 250 words.

Contents

List of Tables

List of Figures

<Note: set up autoreferencing for figures and tables in your document>

1 Research Question and Rationale

As the global population continues to grow and require more food and fuel resources, deforestation trends are expected to accelerate. Deforestation can be problematic as forests provide ecosystem services such as carbon storage, nutrient cycling, water filtration, and wildlife habitat. Agriculture is one of the most commonly sited drivers of deforestation, in addition to being a source of emissions. Given that forest and agriculture can be contending land uses, this research examines the relationship between agriculture and forest as land uses and, more broadly, explores the causes and impacts of changes in land use.

This research looks at the trends and tradeoffs in land use across countries from 1990 – 2014. Primary questions include:

- Is there a relationship between the percentage of land area dedicated to forest versus agricultural in countries?
- Is there a relationship between land uses (agriculture or forest) and levels of CO2, methane, and NO3 emissions?
- Is there a relationship between access to electricity or renewable electricity output and the percentage of land dedicated to forestry versus agriculture?

The research utilizes a data set from the World Bank that has 135 environment-related variables for 264 countries. I have narrowed the environment variable down to 7 that are relevant to land use. Although the full data set dates back to 1960, I have limited to the time scope of the analysis to between 1990 and 2014 because those are the dates for which forest cover data is available.

2 Dataset Information

2.1 Database Information

Data were collected from the World Bank website. More information can be found here: https://data.worldbank.org/topic/environment Data were downloaded as a full data set called "API_6_DS2_en_csv_v2_10518155". The csv file was saved as "WorldBank_Raw_4.8.19". Variables from among the raw data set were selected to be used in the data set analysed. Only 7 of a possible 135 environment-related variables were retained in the processed data. This csv file was saved as ["WorldBank_Processed"; "WorldBank_Spread" or "WB_Spread"].

2.2 Data Content Information

Country: Originally named "Country.Name", the Country variable includes 264 countries.

Year: The Year variable was originally spread horizontally from 1960 – 2018. The years were gathered in to one column named "Year" and formatted as a date.

Forest: The Forest variable was originally named "Forest area (% of land area)". Its data range is 1990 – 2016.

Agriculture: The Agriculture variable was originally named "Agriculture land (% of land area). Its data range is from 1961 – 2018.

Ag.Methane: The Ag.Methane variable was originally named "Agriculture methane emissions (thousand metric tons of CO2 equivalent). Its data range is 1970 – 2008.

Ag.NO2: The Ag.NO2 variable was originally named "Agriculture nitrous oxide emissions (thousand metric tons of CO2 equivalent). Its data range is 1970 – 2008.

CO2Emissions: The CO2Emissions variable was originally named "CO2 emissions (kt). Its data range is 1960 – 2014.

ElectricityAccess: The ElectricityAccess variable was originally named "Access to electricity (% of total population). Data range is 1990 – 2016.

RenewableElectricity: The RenewableElectricity variable was originally named "Renewable electricity output (% of total electricity output). Its data range is 1990 – 2015.

2.3 Naming conventions and file formats

The files are named according to the following convention: `databasename_datatype_details_stage.format` where: Files are named according to the following naming convention: **databasename** refers to the database from where the data originated **details** are additional descriptive details related to the stage of data wrangling **stage**refers to the stage in data management pipelines (e.g., raw or processed) **format** is a non-proprietary file format (e.g., .csv, .txt)

2.4 Additional Information and Support

For more information, please contact the data assembler, **Rebecca Marx** (rebecca.marx@duke.edu)

Variable	Units	Data Strcuture
Cell 1	Cell 2	Cells
Cell 3	Cell 4	Cells

3 Exploratory Data Analysis and Wrangling

Wrangling the data required multiple operations. First, I used the filter function to select only 9 of a possible 135 variables in the data set. Since the data set was arranged with the years going horizontally, I then had to gather the years in to one column, and create a new column called "Level" that contained the values recorded for each variable in each year.

```
World_Bank_Master <-read.csv("../Raw/WorldBank_Raw2_4.8.19.csv")

#Data Subset
World_Bank_Filter <- filter(World_Bank_Master, Indicator.Name == "Forest area (% of land

WorldBank_Gather <- gather(World_Bank_Filter, "Year", "Level", X1960:X2018)

WorldBank_Gather <- select(WorldBank_Gather, -Indicator.Code)

WorldBank_Spread <-  spread(WorldBank_Gather, Indicator.Name, Level)

#Format as character
WorldBank_Spread$Year <- as.character(WorldBank_Spread$Year)

#create string
WB_String <- substr(WorldBank_Spread$Year, 2, 5)

#Get rid of X in date
WorldBank_Spread$Year = WB_String

#Format as date
#WB_Fixed$Year <- as.Date(WB_Fixed$Year)
WorldBank_Spread$Year <- as.Date(WorldBank_Spread$Year, format = "%Y") #can I get it to

#Change column names
names(WorldBank_Spread) <- c("Country", "Indicator.Code", "Year", "ElectricityAccess", "

#Save processed file
#write.csv(WorldBank_Spread, row.names = FALSE, file = "../Processed/WorldBank_Process

Five_Countries <- filter(WorldBank_Spread, Country == "Brazil" | Country == "Kenya" | Co

WB_Spread <- WorldBank_Spread %>%
  na.exclude #check if I need this for my tests

WB_Brazil <- filter(WB_Spread, Country == "Brazil")
```

```
<Include R chunks for 5+ lines of summary code (display code and output), 3+ exploratory
```

After the data was wrangled into a format that would function well with my intended R operations, I was able to begin exploring the data set.

```r
#5+ lines of summary

colnames(WorldBank_Spread)
```

```
##  [1] "Country"               "Indicator.Code"        "Year"
##  [4] "ElectricityAccess"     "Agriculture"           "Ag.Methane"
##  [7] "Ag.NO2"                "CO2Emissions"          "Forest"
## [10] "RenewableElectricity"
```

```r
class(WorldBank_Spread$Year)
```

```
## [1] "Date"
```

```r
dim(WorldBank_Spread)
```

```
## [1] 15576    10
```

```r
head(WorldBank_Spread)
```

```
##       Country Indicator.Code       Year ElectricityAccess Agriculture
## 1 Afghanistan            AFG 1960-04-15                NA          NA
## 2 Afghanistan            AFG 1961-04-15                NA    57.74592
## 3 Afghanistan            AFG 1962-04-15                NA    57.83782
## 4 Afghanistan            AFG 1963-04-15                NA    57.91441
## 5 Afghanistan            AFG 1964-04-15                NA    58.01091
## 6 Afghanistan            AFG 1965-04-15                NA    58.01397
##   Ag.Methane Ag.NO2 CO2Emissions Forest RenewableElectricity
## 1         NA     NA      414.371     NA                   NA
## 2         NA     NA      491.378     NA                   NA
## 3         NA     NA      689.396     NA                   NA
## 4         NA     NA      707.731     NA                   NA
## 5         NA     NA      839.743     NA                   NA
## 6         NA     NA     1008.425     NA                   NA
```

```r
summary(WorldBank_Spread)
```

```
##             Country      Indicator.Code       Year
##  Afghanistan    :   59   ABW    :   59   Min.   :1960-04-15
##  Albania        :   59   AFG    :   59   1st Qu.:1974-04-15
##  Algeria        :   59   AGO    :   59   Median :1989-04-15
##  American Samoa:   59   ALB    :   59   Mean   :1989-04-14
##  Andorra        :   59   AND    :   59   3rd Qu.:2004-04-15
##  Angola         :   59   ARB    :   59   Max.   :2018-04-15
```

```
##  (Other)        :15222   (Other):15222
##  ElectricityAccess  Agriculture        Ag.Methane          Ag.NO2
##  Min.   :  0.00    Min.   : 0.2628   Min.    :       0   Min.    :      0.0
##  1st Qu.: 53.11    1st Qu.:20.5547   1st Qu.:     120   1st Qu.:     86.9
##  Median : 93.94    Median :37.3659   Median :    3300   Median :   2302.9
##  Mean   : 75.04    Mean   :37.0790   Mean    : 117609   Mean    :  63590.8
##  3rd Qu.:100.00    3rd Qu.:52.3930   3rd Qu.:  24198   3rd Qu.:  15076.6
##  Max.   :100.00    Max.   :93.4407   Max.    :3464398   Max.    :2242932.7
##  NA's   :8618      NA's   :2521      NA's    :5056      NA's    :5056
##   CO2Emissions          Forest          RenewableElectricity
##  Min.   :      -81   Min.   :    0.00   Min.   :  0.000
##  1st Qu.:      964   1st Qu.:   12.50   1st Qu.:  0.465
##  Median :    11463   Median :   31.18   Median : 16.961
##  Mean   :   736069   Mean   :   42.70   Mean   : 28.211
##  3rd Qu.:   143107   3rd Qu.:   46.96   3rd Qu.: 49.255
##  Max.   :36138285   Max.   :16735.00   Max.   :100.000
##  NA's   :3321       NA's   :8717       NA's   :8738
```

```
summary(WorldBank_Spread$Agriculture)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.2628 20.5547 37.3659 37.0790 52.3930 93.4407    2521
```

```
summary(WorldBank_Spread$Forest)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##     0.00    12.50    31.18    42.70    46.96 16735.00     8717
```

```
summary(WorldBank_Spread$RenewableElectricty)
```

```
## Length  Class   Mode
##      0   NULL   NULL
```

```
summary(WB_Spread$Ag.Methane)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0     687    4638  138432   36696 3464398
```

Figure1 plots agriculture and forest land covers to see if there is a noticable trend. It appears that there is a tendency for forest cover to decrease as agriculture cover increases.

```
## function (expr)
## {
##     enexpr(expr)
## }
## <bytecode: 0x000000001a022e28>
## <environment: namespace:rlang>
```

A second exploratory graph (Figure2) plots forest levels for all 264 countries over time. The

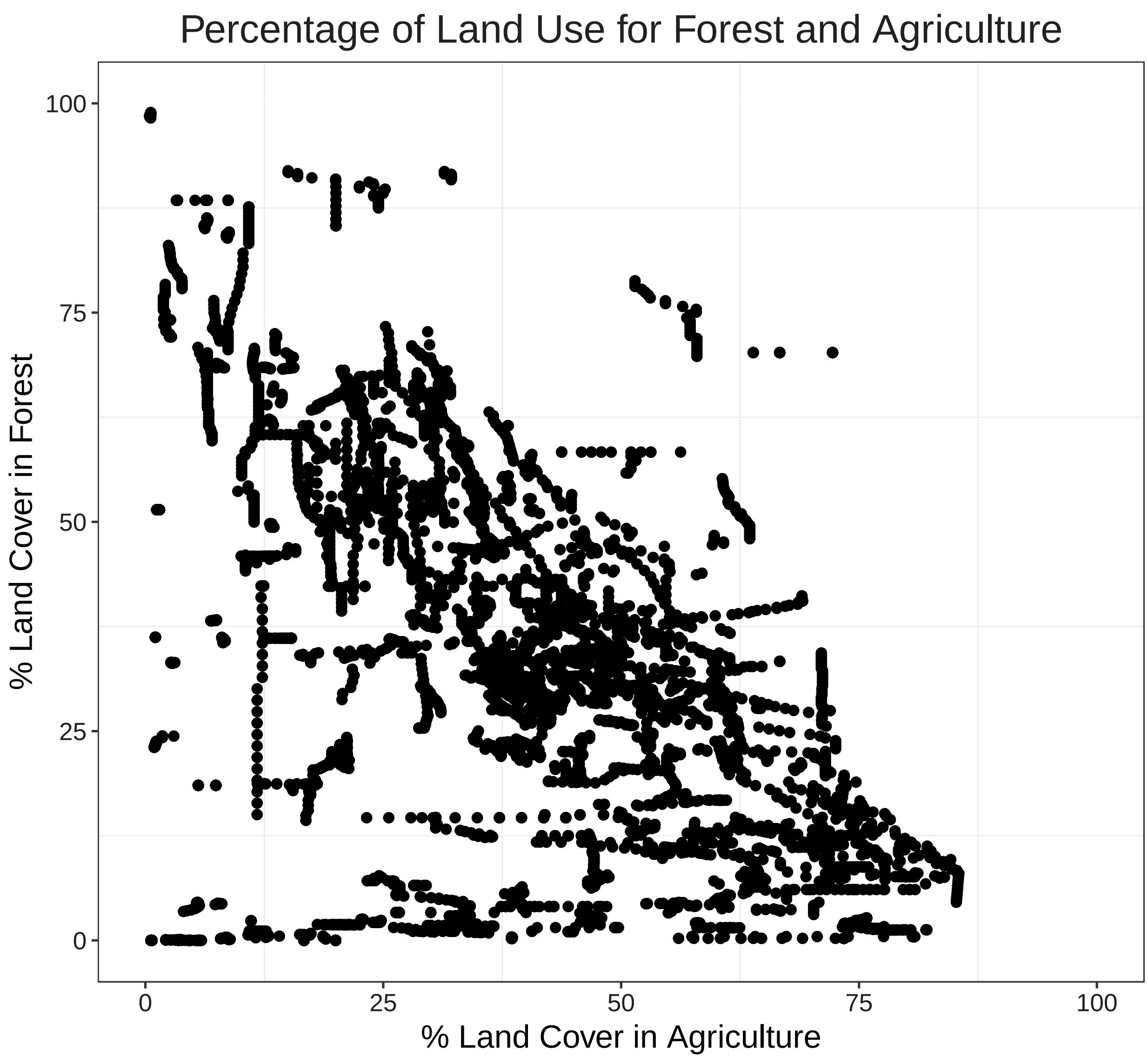


Figure 1: Percentage of Land Use for Forest and Agriculture

resulting graph shows that there is some movement over time, but many countries see constant levels of land use. Similarly, the plot of agriculture over time shows some movement over time, but many countries see constant levels of land use for agriculture. In later graphs I decided to reduce the number of countries included to be able to better visualized changes.

```
## function (expr)
## {
##     enexpr(expr)
## }
## <bytecode: 0x000000001a022e28>
## <environment: namespace:rlang>
```

Since the research attempts to decipher whether a change in forest levels is related to a change in agriculture levels, I also wanted to see if a change in the amount of land dedicated to agriculture leads to a change in various emissions levels. The following graphs explore these relationships using the 5 selected countries.

```
## function (expr)
## {
##     enexpr(expr)
## }
## <bytecode: 0x000000001a022e28>
## <environment: namespace:rlang>
```

"' Deforestation is also associated with the release of Carbon, so I checked to see whether a relationship between forst and carbon emissions could be seen.

```
ForestVCarbon <-
  ggplot(WorldBank_Spread) +
  geom_point(aes(x = Forest, y = CO2Emissions)) +
  ggtitle("Forest vs. Carbon") +
  ylab(expression("Carbon")) +
  xlab(expression("% Land in Forest")) +
  scale_x_continuous(limits = c(0,100))
print(ForestVCarbon)
```

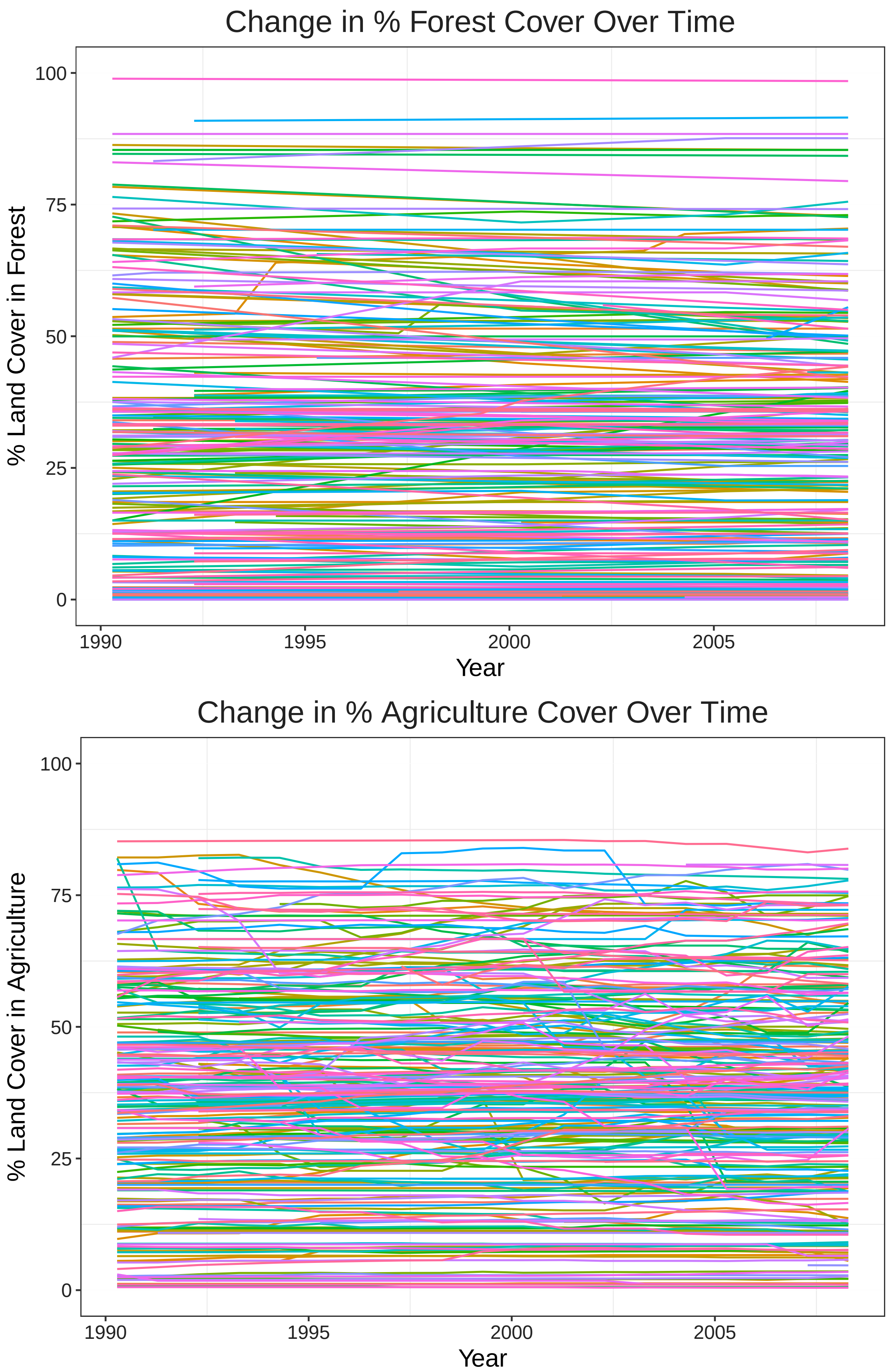


Figure 2: Percentage of Land Use for Forest and Agriculture

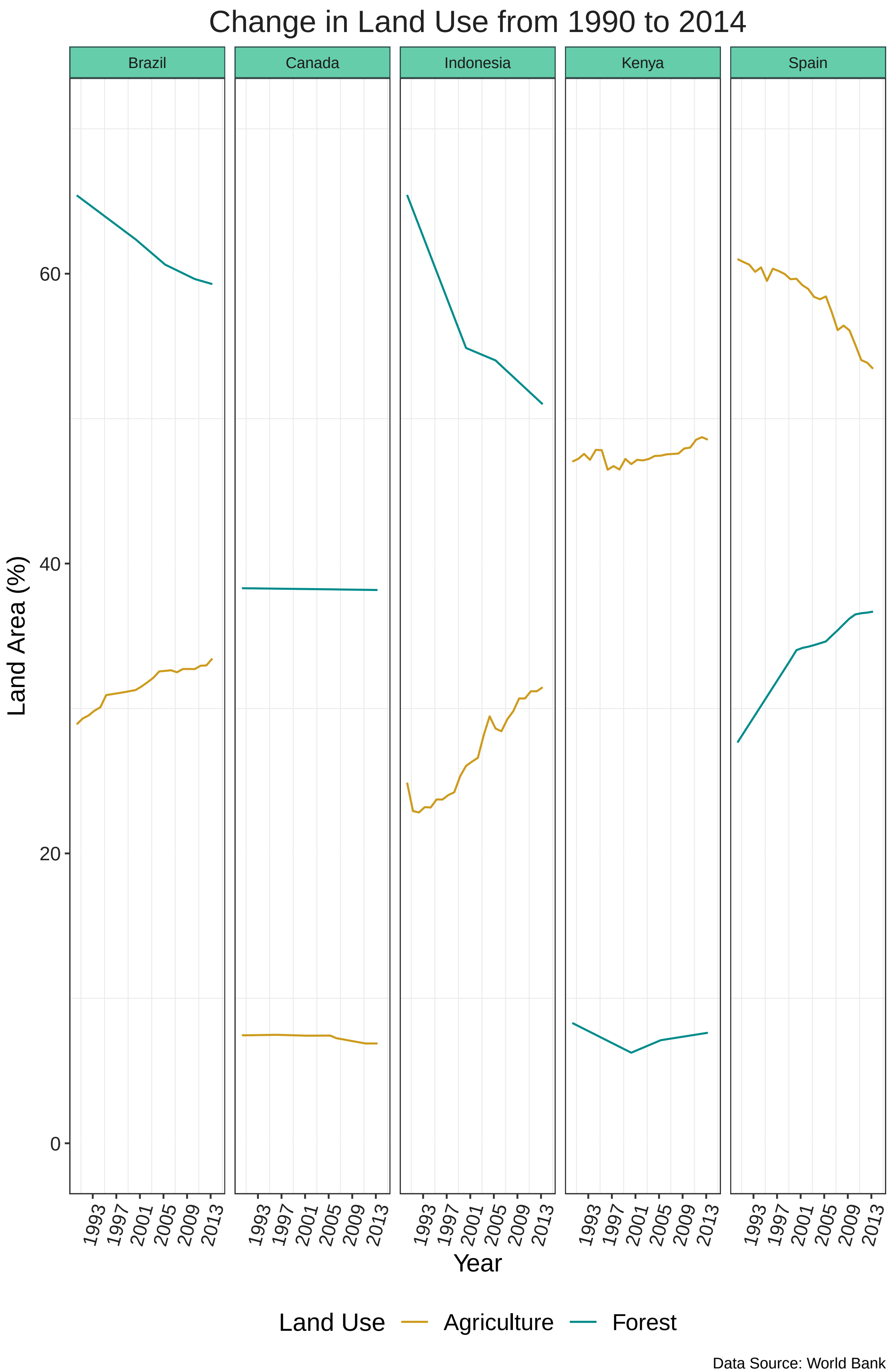


Figure 3: Change in Land Use from 1990 to 2014 for Five Countries

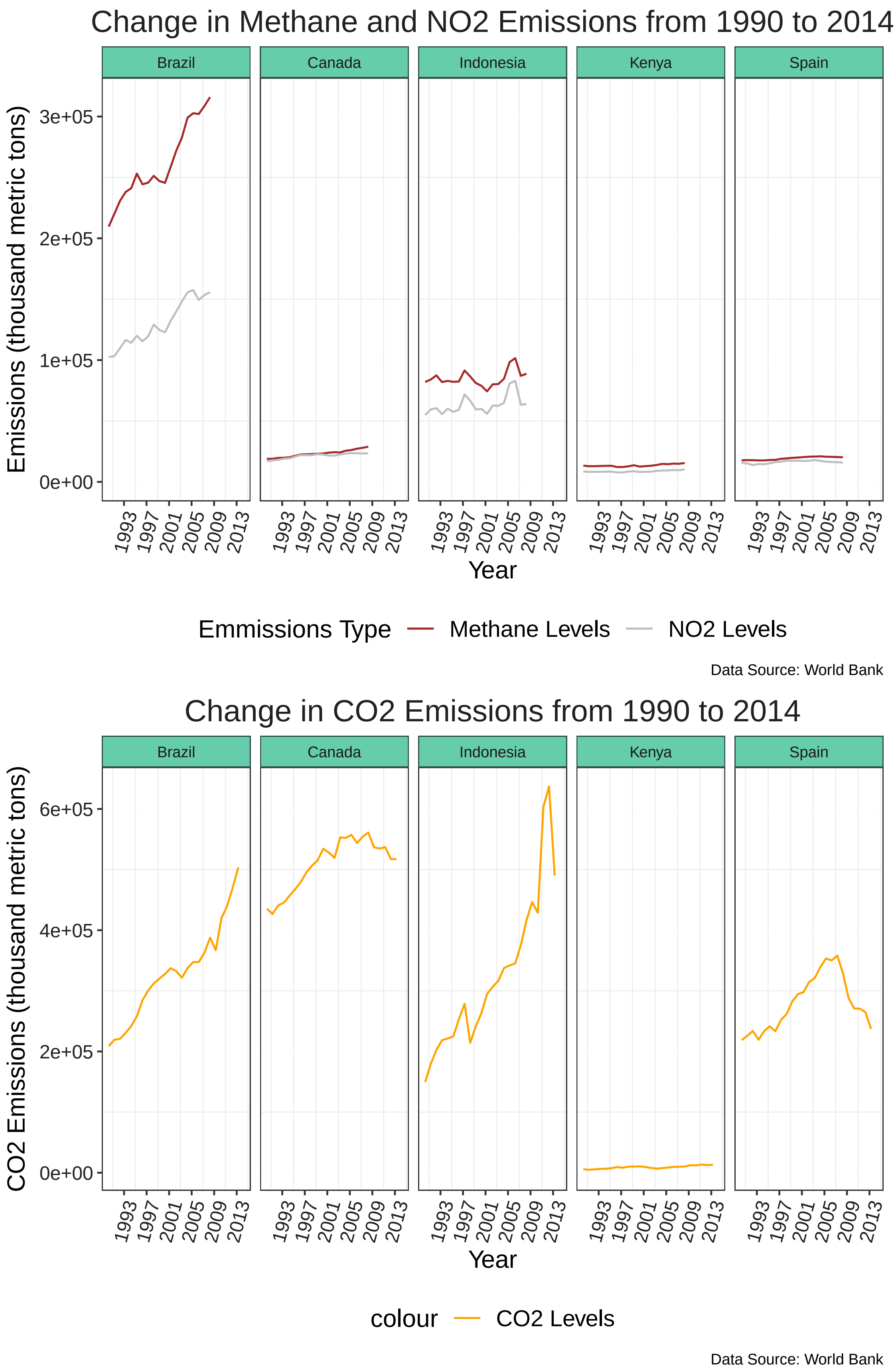


Figure 4: Change in Emissions Use from 1990 to 2014 for Five Countries

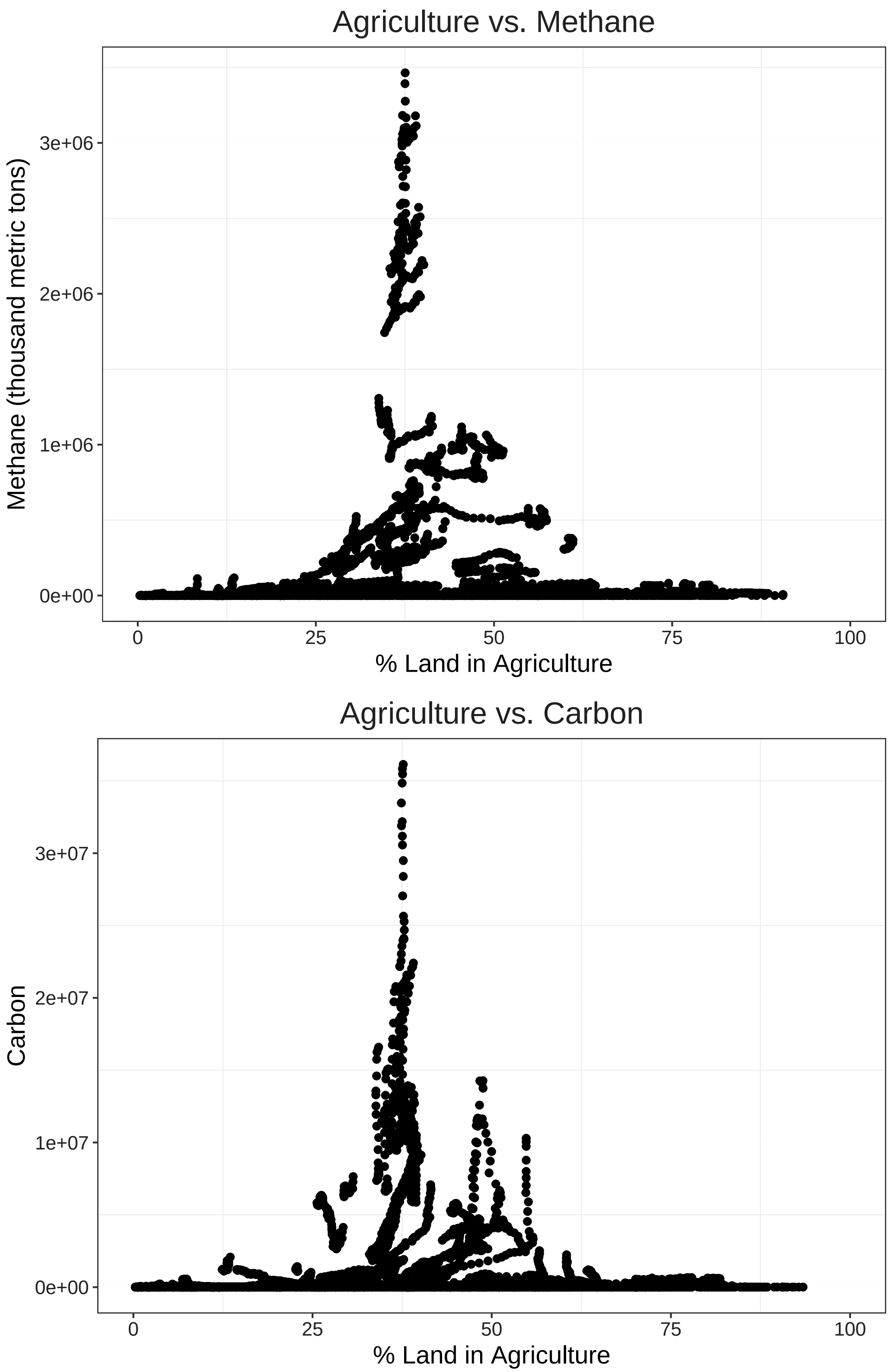
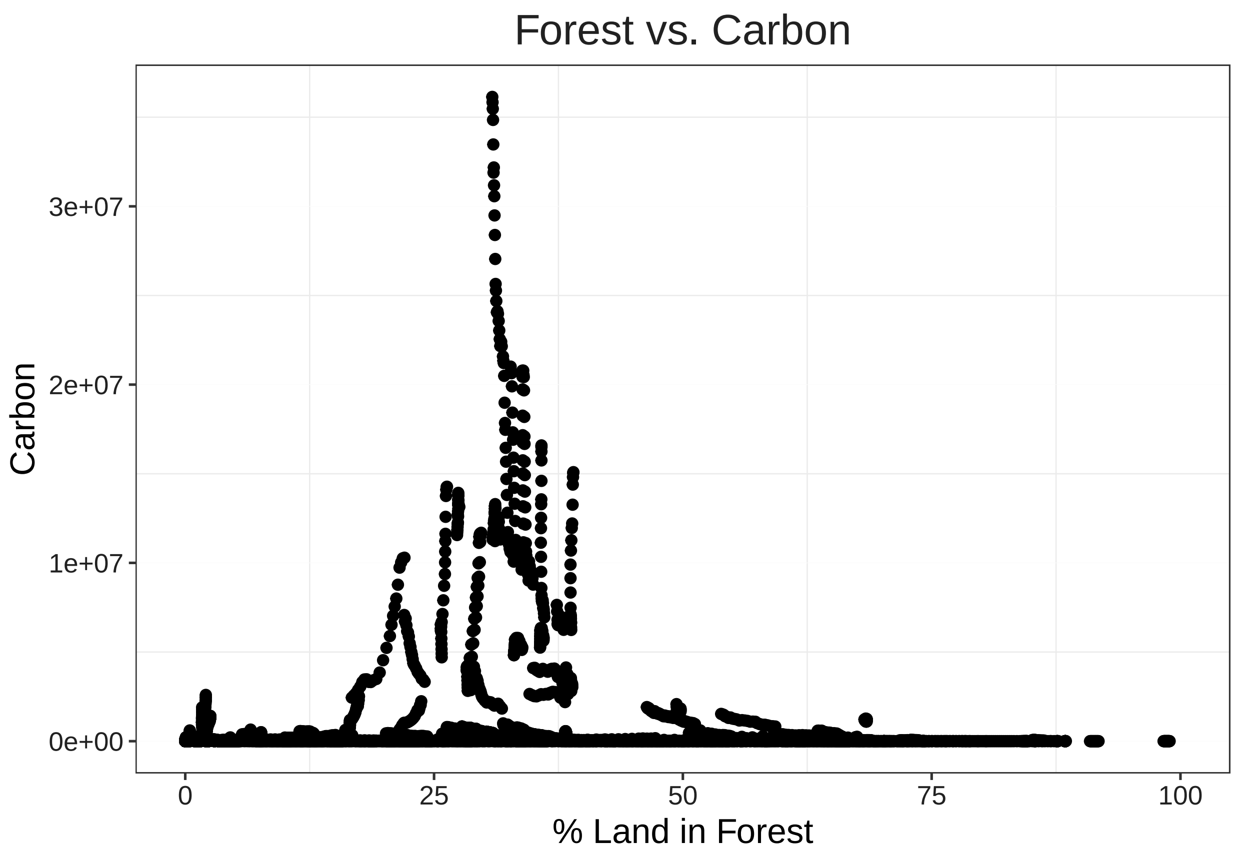
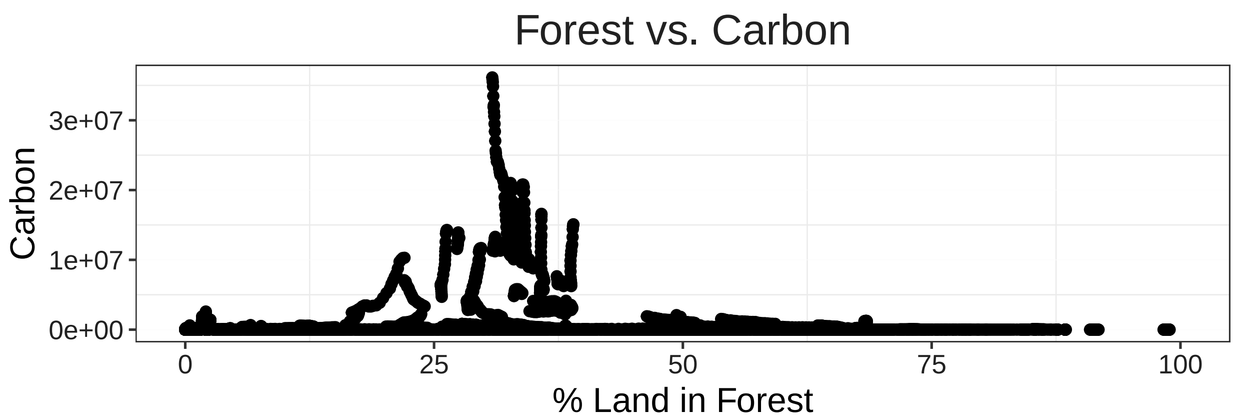
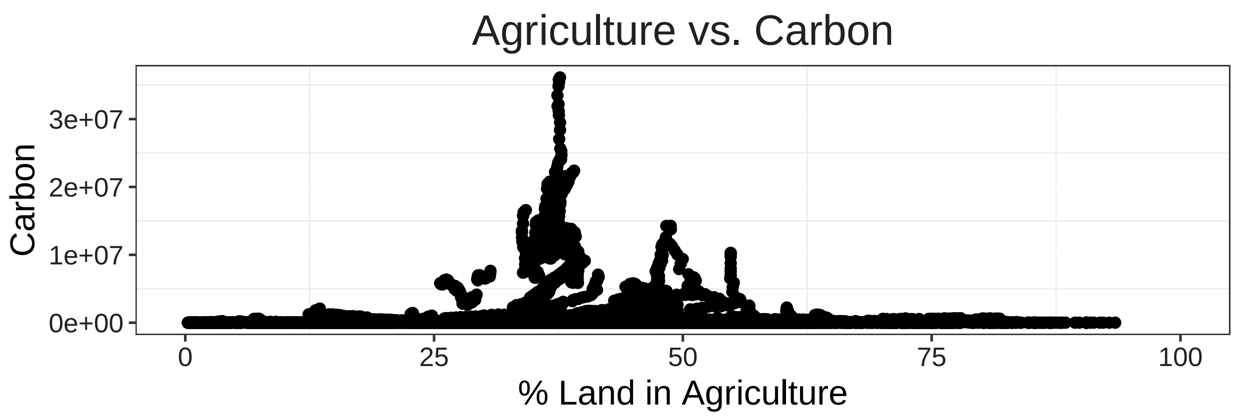


Figure 5: Emissions versus Agriculture



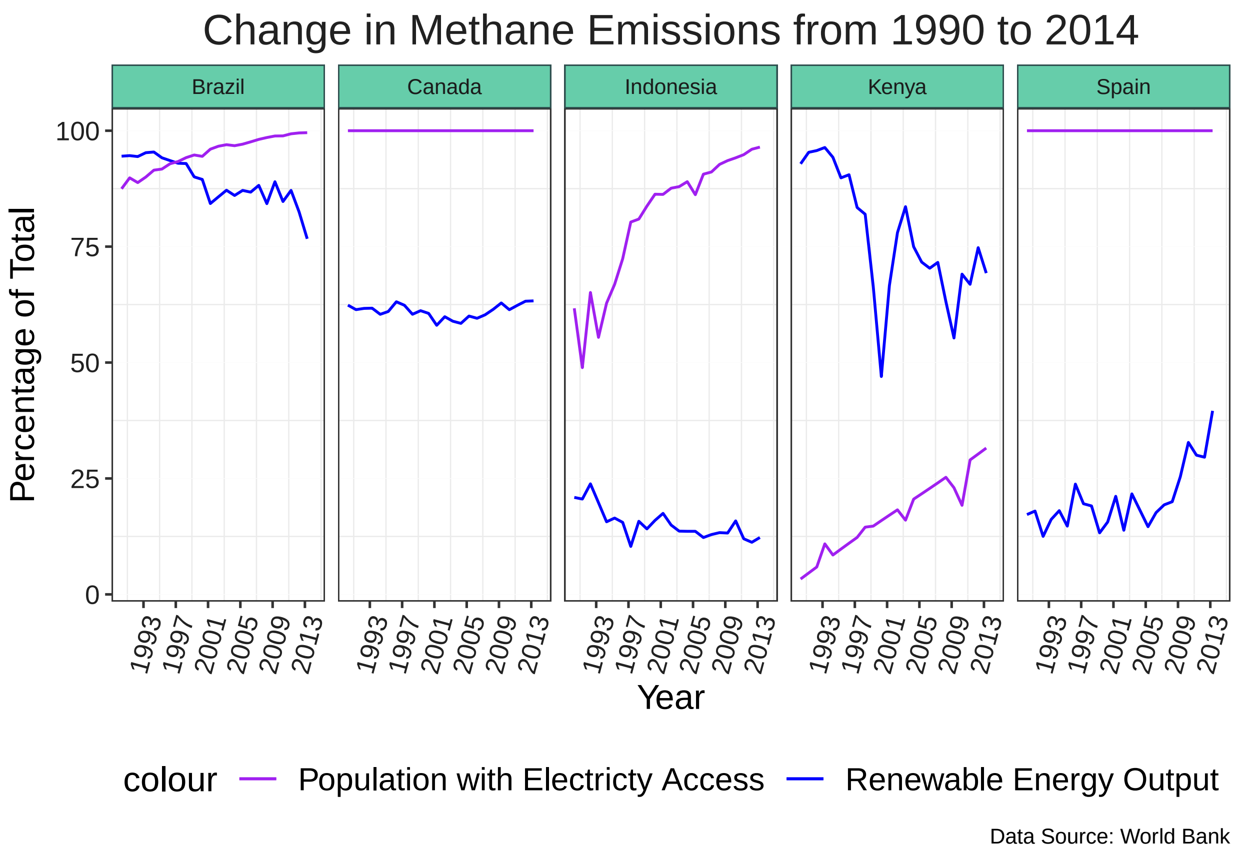

```
grid.arrange(ForestVCarbon, AgVCarbon)
```

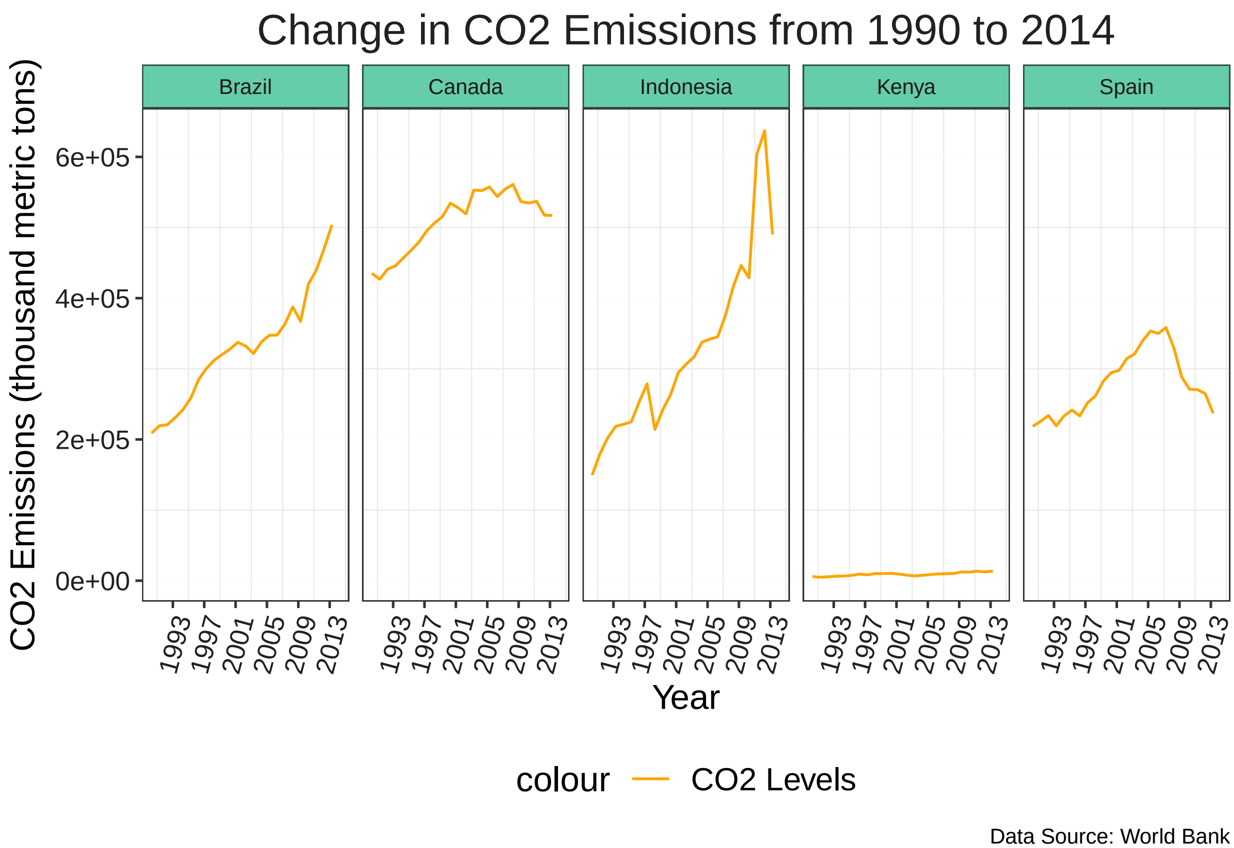


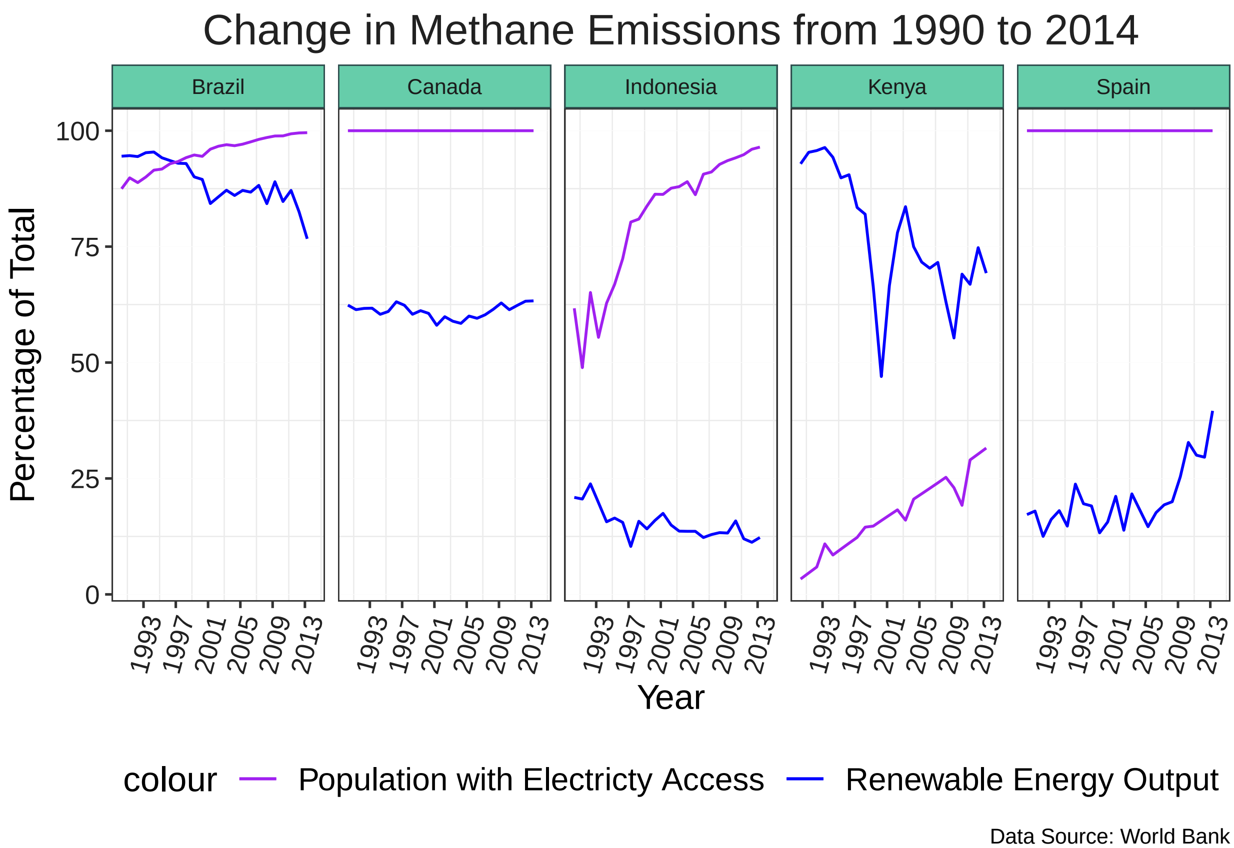


“‘

```
FiveCountries.RE <-
  ggplot(Five_Countries) +
  geom_line(aes(x = Year, y = RenewableElectricity, color = "Renewable Energy Output"))
  geom_line(aes(x = Year, y = ElectricityAccess, color = "Population with Electricty Acc
  facet_grid(cols = vars(Country)) +
  ggtitle("Change in Methane Emissions from 1990 to 2014") +
  ylab(expression("Percentage of Total")) +
  scale_x_date(limits = as.Date(c("1990-04-09", "2014-04-09")),
  date_breaks = "48 months", date_labels = "%Y") +
  theme(axis.text.x = element_text(angle = 75, hjust = 1)) +
  scale_color_manual(values = c("purple", "blue")) +
  labs(caption = "Data Source: World Bank") +
  theme(strip.background = element_rect(fill= "aquamarine3", "darkslategray"))
print(FiveCountries.RE)
```


Change in Methane Emissions from 1990 to 2014
Brazil
Canada
Indonesia
Kenya
Spain
Percentage of Total
100
75
50
25
0
1993
1997
2001
2005
2009
2013
Year
colour
Population with Electricty Access
Renewable Energy Output
Data Source: World Bank


Change in CO2 Emissions from 1990 to 2014
Brazil
Canada
Indonesia
Kenya
Spain
CO2 Emissions (thousand metric tons)
6e+05
4e+05
2e+05
0e+00
1993
1997
2001
2005
2009
2013
Year
colour
CO2 Levels
Data Source: World Bank


Change in Methane Emissions from 1990 to 2014
Brazil
Canada
Indonesia
Kenya
Spain
Percentage of Total
100
75
50
25
0
1993
1997
2001
2005
2009
2013
Year
colour
Population with Electricty Access
Renewable Energy Output
Data Source: World Bank

4 Analysis

```r
#Statistical Test 1: How has forest changed over time
Forest.Time1 <- gls(data = WB_Spread,
                    Forest ~ Year,
                    method = "REML")
summary(Forest.Time1)
```

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year
##   Data: WB_Spread
##        AIC      BIC    logLik
##   49977.96 49997.13 -24985.98
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 83.48146  5.839645 14.295639       0
## Year        -0.00425  0.000533 -7.987907       0
##
##  Correlation:
##      (Intr)
## Year -0.984
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.74247841 -0.33743857 -0.09641382  0.15552883 16.46833243
##
## Residual standard error: 69.98862
## Degrees of freedom: 4408 total; 4406 residual
```

```r
Forest.Time2 <- lme(data = WB_Spread,
                Forest ~ Year,
                random = ~1| Country)
summary(Forest.Time2)
```

```
## Linear mixed-effects model fit by REML
##  Data: WB_Spread
##        AIC      BIC    logLik
##   49431.95 49457.52 -24711.98
##
## Random effects:
##  Formula: ~1 | Country
##         (Intercept) Residual
## StdDev:    30.52692 62.84915
##
```

```
## Fixed effects: Forest ~ Year
##                 Value Std.Error   DF   t-value p-value
## (Intercept) 82.96649  5.653843 4163 14.674352       0
## Year        -0.00421  0.000482 4163 -8.743567       0
##  Correlation:
##      (Intr)
## Year -0.923
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.55113019 -0.14187377 -0.01873545  0.09332491 16.07509035
##
## Number of Observations: 4408
## Number of Groups: 244
```

A fixed effects model was used to see how land percetnage of forest area accross the whoel data set changes over time. The results show that, on average, forest decreases by -.00485% each year. Was not worride about seasonality of the data because there is only one data point per year. As such, I did not calculate autocorrelations. I did add country as a random effect.

```
For.Ag.1 <- gls(data = WB_Spread,
                    Forest ~ Year + Agriculture,
                    method = "REML")
summary(For.Ag.1)
```

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year + Agriculture
##   Data: WB_Spread
##        AIC      BIC    logLik
##   49908.57 49934.13 -24950.29
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 100.30274  6.103018 16.434939       0
## Year         -0.00423  0.000528 -8.007157       0
## Agriculture  -0.44673  0.051202 -8.724865       0
##
##  Correlation:
##             (Intr) Year
## Year        -0.931
## Agriculture -0.316 -0.006
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.98942500 -0.26448939 -0.05947803  0.14053065 16.59420967
##
```

```
## Residual standard error: 69.39948
## Degrees of freedom: 4408 total; 4405 residual
```

```
For.Ag.2 <-  lme(data = WB_Spread,
              Forest ~ Year + Agriculture,
              random = ~1| Country)
summary(For.Ag.2)
```

```
## Linear mixed-effects model fit by REML
##  Data: WB_Spread
##       AIC      BIC    logLik
##   49413.7 49445.65 -24701.85
##
## Random effects:
##  Formula: ~1 | Country
##         (Intercept) Residual
## StdDev:    29.32764 62.80075
##
## Fixed effects: Forest ~ Year + Agriculture
##                 Value Std.Error   DF   t-value p-value
## (Intercept) 101.78896  6.830290 4162 14.902583       0
## Year         -0.00421  0.000481 4162 -8.754338       0
## Agriculture  -0.49055  0.101138 4162 -4.850333       0
##  Correlation:
##             (Intr) Year
## Year        -0.763
## Agriculture -0.568  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.53066671 -0.13903414 -0.01968184  0.09295492 16.10850732
##
## Number of Observations: 4408
## Number of Groups: 244
```

```
#anova(Forest.Fixed, Test2) # Said: fitted objects with different fixed effects. REML
```

Results of first set of tests:

I am interested in whether access to electricity or the levels of renewable energy produced in countries have a relationship with deforestation. This is becasue a lack of electricity could drive the use of forest biomass products as an energy source. Electricity access appears to have a negative impact on forest (-.333) which may be because electricity is needed for some agricultural operations, or electricity is being produced via biofuels, the production of which might require more forest to be converted to agricultural land. (3)

With that theory in mind, I was interested in whether there might be a relationship between forested land and renewable energy because energy produced from biomass is considered to be renewable. However, the result of my gls revealed that a 1 unit increase in renewable energy is associated with a .18% increase in forest land, meaning the production of biofule is likely not a main driver of deforestation.

```
#Looking at Electricity Accesss and Renewable Energy as a Drive of the Ag. / Forest Tr

#Add Electricity
Forest.Ag.Elec <- gls(data = WB_Spread,
                    Forest ~ Year + Agriculture + ElectricityAccess,
                    method = "REML")
Forest.Ag.Elec
```

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year + Agriculture + ElectricityAccess
##   Data: WB_Spread
##   Log-restricted-likelihood: -24906.59
##
## Coefficients:
##       (Intercept)              Year       Agriculture ElectricityAccess
##     117.273716122      -0.003583843      -0.514782445      -0.299412308
##
## Degrees of freedom: 4408 total; 4404 residual
## Residual standard error: 68.68236
```

```
#Electricity only
Forest.Elec <- gls(data = WB_Spread,
                    Forest ~ Year + ElectricityAccess,
                    method = "REML")
Forest.Elec
```

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year + ElectricityAccess
##   Data: WB_Spread
##   Log-restricted-likelihood: -24954.6
##
## Coefficients:
##       (Intercept)              Year ElectricityAccess
##      95.826798295      -0.003705033      -0.256536944
##
## Degrees of freedom: 4408 total; 4405 residual
## Residual standard error: 69.4595
```

```
Forest.RE <- gls(data = WB_Spread,
                    Forest ~ Year + RenewableElectricity,
                    method = "REML")
Forest.RE
```

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year + RenewableElectricity
##   Data: WB_Spread
##   Log-restricted-likelihood: -24969.59
##
## Coefficients:
##          (Intercept)                 Year RenewableElectricity
##          75.893689686         -0.004120715          0.199889723
##
## Degrees of freedom: 4408 total; 4405 residual
## Residual standard error: 69.69694
```

```
#On that note,

# Regression w/o time
Reg1 <- lm(Forest ~ Agriculture, WorldBank_Spread)
Reg1
```

```
##
## Call:
## lm(formula = Forest ~ Agriculture, data = WorldBank_Spread)
##
## Coefficients:
## (Intercept)  Agriculture
##     52.5762      -0.4444
```

```
#A 1 unit increas in agriculture leads to a -.44 decrease in forest
```

As a second statistical test I used Pettit's test, a nonparametric test that determines if there is a shift in the central tendency of the time series. It also determines at what point in time the changepoint occurs. I used this to see if there was a point in time in particular where these seemed to be a noticeable change in forest cover. I applied the test to Brazil which graphically appears to have change points to see if 1) there was a change poing for agriculture and forest and 2) to see if the change points occured in the same year for each.

```
#Statistical Test 2: Pettitts Test: Looking at Change Points in the Data

#Statistical Test 2: Any change points for full forest data?
pettitt.test(WB_Spread$Forest)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Spread$Forest
## U* = 463360, p-value = 5.895e-07
## alternative hypothesis: two.sided
## sample estimates:
```

```
## probable change point at time K
##                            1739
```

```r
#Probable change point at time 1428 which doesn't exist.


#Changes in Brazil Forest
pettitt.test(WB_Brazil$Forest)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Brazil$Forest
## U* = 90, p-value = 0.002386
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K                            <NA>
##                               9                              10
```

```r
#9 (of 19): 1998; p-value .048 (significant at the .05 level)

pettitt.test(WB_Brazil$Forest[9:19])
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Brazil$Forest[9:19]
## U* = 30, p-value = 0.04852
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K                            <NA>
##                               5                               6
```

```r
#Changes in Brazil Agriculture
#5 (9+5 = 14: 2003); p-value .0023852 (significant at the .01 level)

pettitt.test(WB_Brazil$Forest[14:19])
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Brazil$Forest[14:19]
## U* = 9, p-value = 0.2907
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                               3
```

```r
#3 14 + 3 = 17: 2006 (p-value .29: not significant)

pettitt.test(WB_Brazil$Agriculture)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Brazil$Agriculture
## U* = 90, p-value = 0.002386
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K                               <NA>
##                               9                                 10
```

```r
#9 (of 19): 1998; p-value .002386 (significant at the .01 level)

pettitt.test(WB_Brazil$Agriculture[9:19])
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Brazil$Agriculture[9:19]
## U* = 30, p-value = 0.04852
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K                               <NA>
##                               5                                  6
```

```r
#5 (9+5 = 14: 2003); p-value .04852 (significant at the .05 level

pettitt.test(WB_Brazil$Agriculture[14:19])
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Brazil$Agriculture[14:19]
## U* = 6, p-value = 0.8487
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                               2
```

```r
#2 (14 +2 = 16) p-value .8487 (not significant)
```

Pettitt's applied to a single country (Brazil) initially detects a change point in Forest and Agriculture in the same year. The change for both was in place 9, which is the year 1998.

The significant change points for agriculture and forest in Brazil occured in the same years (1998 and 2003), furhter supporting the notion that there is a relationship between the two.

[Make a graph of Brazil that shows the change points]

The next series of tests address the question: Is there a relationship between land uses (agriculture or forest) and levels of CO2, methane, and NO3 emissions?

```
#Statistical Test 3.1: Agriculture & Methane Emmissions

AgMethane <- gls(data = WB_Spread,
                 Ag.Methane ~ Year + Agriculture,
                 method = "REML")
AgMethane
```

```
## Generalized least squares fit by REML
##   Model: Ag.Methane ~ Year + Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -63169.07
##
## Coefficients:
##  (Intercept)         Year  Agriculture
## 1.172434e+05 3.206730e-02 5.445887e+02
##
## Degrees of freedom: 4408 total; 4405 residual
## Residual standard error: 406812.4
```

```
#Statistica Test 3.2: Forest & Methane Emmissions
ForestMethane <- gls(data = WB_Spread,
                  Ag.Methane ~ Year + Forest,
                  method = "REML")
summary(ForestMethane)
```

```
## Generalized least squares fit by REML
##   Model: Ag.Methane ~ Year + Forest
##   Data: WB_Spread
##        AIC      BIC    logLik
##   126331.7 126357.3 -63161.87
##
## Coefficients:
##                 Value Std.Error  t-value p-value
## (Intercept) 104949.59  34655.09 3.028404  0.0025
## Year             1.73      3.11 0.557427  0.5773
## Forest         392.90     87.40 4.495392  0.0000
##
##  Correlation:
##        (Intr) Year
```

```
## Year   -0.980
## Forest -0.211  0.119
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -0.9349863 -0.3368465 -0.3079788 -0.2493750  8.1856088
##
## Residual standard error: 406034.1
## Degrees of freedom: 4408 total; 4405 residual
```

```
#Statistical Test 3.3: Interaction of Forest and Agriculture on Methane Emissions #The
Int.Methane <-  gls(data = WB_Spread,
                    Ag.Methane ~ Year + Forest * Agriculture,
                    method = "REML")
Int.Methane
```

```
## Generalized least squares fit by REML
##   Model: Ag.Methane ~ Year + Forest * Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -63123.24
##
## Coefficients:
##        (Intercept)               Year             Forest
##      170978.276563           2.186635       -2547.247815
##        Agriculture Forest:Agriculture
##       -1627.888779          81.435201
##
## Degrees of freedom: 4408 total; 4403 residual
## Residual standard error: 403492.6
```

As the percetage of land under agriculture increases by 1, methane emissions increase by 6.57 thousand metric tons.

As Forest land increases by 1%, methane emissions increase by .035 thoushand metric tons (or 35 metirc tons), meaning that deforestation is not strongly related to methan emissions.

However, if the interaction of forest and agriculture land cover is considered, an increase of 1 leads to an increase of 98.86 thousand metric tons of methane.

```
#Carbon Emissions Tests

#Test 4.1: Forest and Carbon
Forest.CO2 <- gls(data = WB_Spread,
                  CO2Emissions ~ Year + Forest,
                  method = "REML")
Forest.CO2
```

```
## Generalized least squares fit by REML
```

```
##   Model: CO2Emissions ~ Year + Forest
##   Data: WB_Spread
##   Log-restricted-likelihood: -71822.3
##
## Coefficients:
##  (Intercept)         Year       Forest
## 311972.83203     54.86307     30.35953
##
## Degrees of freedom: 4408 total; 4405 residual
## Residual standard error: 2900052
```

```
#Test 4.2: Agriculture and Carbon
Ag.CO2 <- gls(data = WB_Spread,
                 CO2Emissions ~ Year + Agriculture,
                 method = "REML")
Ag.CO2
```

```
## Generalized least squares fit by REML
##   Model: CO2Emissions ~ Year + Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -71820.95
##
## Coefficients:
##  (Intercept)         Year  Agriculture
## 274157.42618     54.67258   1071.59529
##
## Degrees of freedom: 4408 total; 4405 residual
## Residual standard error: 2899970
```

```
#Test 4.3: Interaction of Forest & Agriculture on Carbon
Int.CO2 <- gls(data = WB_Spread,
                   CO2Emissions ~ Year + Forest * Agriculture,
                   method = "REML")
Int.CO2
```

```
## Generalized least squares fit by REML
##   Model: CO2Emissions ~ Year + Forest * Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -71789.87
##
## Coefficients:
##        (Intercept)               Year             Forest
##       847291.77408           57.25021       -17883.04480
##        Agriculture Forest:Agriculture
##       -13175.94334          492.65459
##
```

```
## Degrees of freedom: 4408 total; 4403 residual
## Residual standard error: 2888536
```

1% increase in forest leads to a -235.25 decrease in CO2. 1% increase in Ag. leads to an increase of 1493 kt of carbon. Interaction: How do you interpret?

5 Summary and Conclusions