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 9 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

3 Exploratory Data Analysis and Wrangling

```
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

class(WorldBank_Spread$Year)
```

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

```
#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

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

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

```r
#5+ lines of summary

colnames(WorldBank_Spread)
```

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

```r
dim(WorldBank_Spread)
```

```
## [1] 15576    12
```

```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 Aquaculture ArableLand CO2Emissions Forest
## 1         NA     NA          NA         NA      414.371     NA
## 2         NA     NA          NA   11.71767      491.378     NA
## 3         NA     NA          NA   11.79426      689.396     NA
## 4         NA     NA          NA   11.87085      707.731     NA
## 5         NA     NA          NA   11.94743      839.743     NA
## 6         NA     NA          NA   11.94743     1008.425     NA
##   RenewableElectricity
## 1                   NA
## 2                   NA
## 3                   NA
## 4                   NA
## 5                   NA
## 6                   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
##   Aquaculture          ArableLand       CO2Emissions
##  Min.   :        0   Min.   : 0.0012   Min.   :     -81
##  1st Qu.:       68   1st Qu.: 3.5315   1st Qu.:     964
##  Median :     3758   Median : 9.5558   Median :   11463
##  Mean   :  1601961   Mean   :13.1413   Mean   :  736069
##  3rd Qu.:    95447   3rd Qu.:17.5690   3rd Qu.:  143107
##  Max.   :106004184   Max.   :73.3886   Max.   :36138285
##  NA's   :4696        NA's   :2658      NA's   :3321
##      Forest          RenewableElectricity
##  Min.   :    0.00   Min.   :  0.000
##  1st Qu.:   12.50   1st Qu.:  0.465
##  Median :   31.18   Median : 16.961
##  Mean   :   42.70   Mean   : 28.211
##  3rd Qu.:   46.96   3rd Qu.: 49.255
##  Max.   :16735.00   Max.   :100.000
##  NA's   :8717       NA's   :8738
```

```r
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
```

```r
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
```

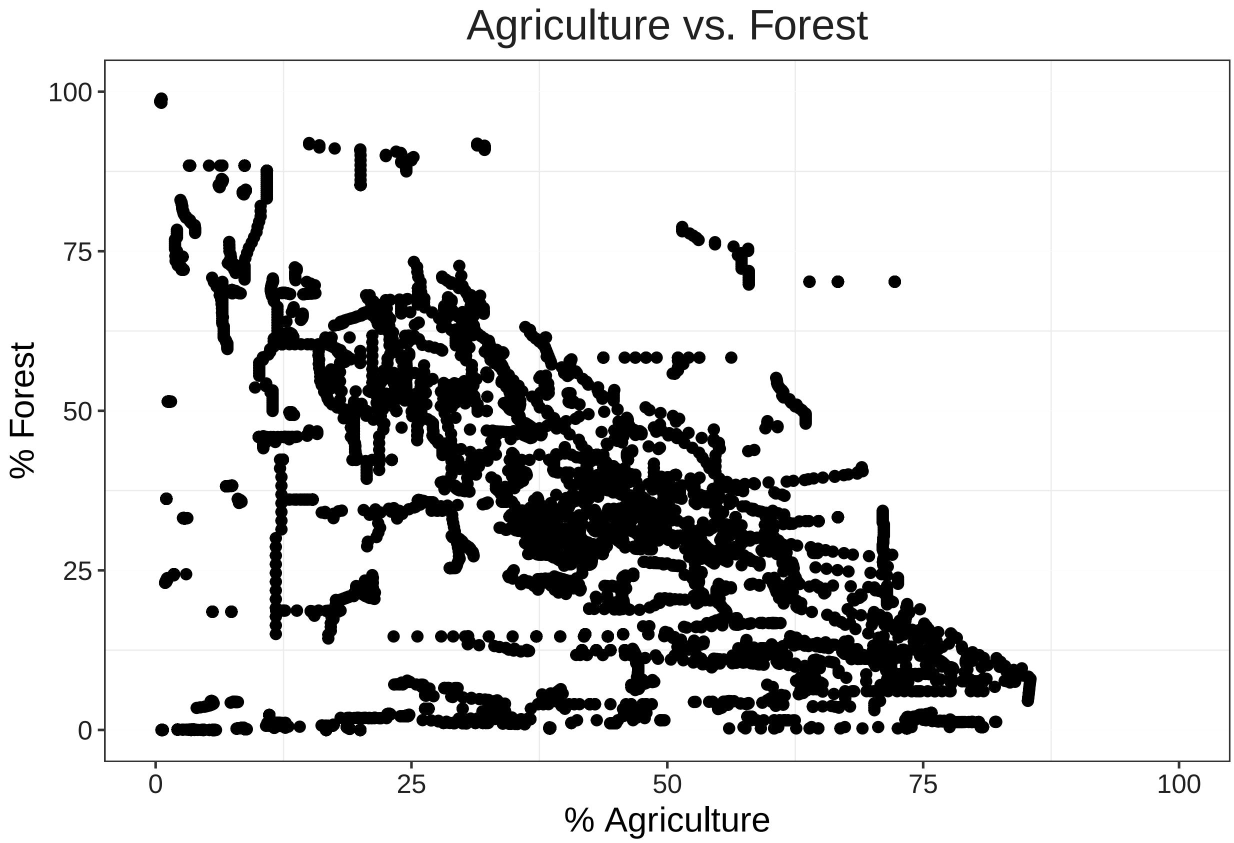
```r
summary(WorldBank_Spread$RenewableElectricty)
```

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

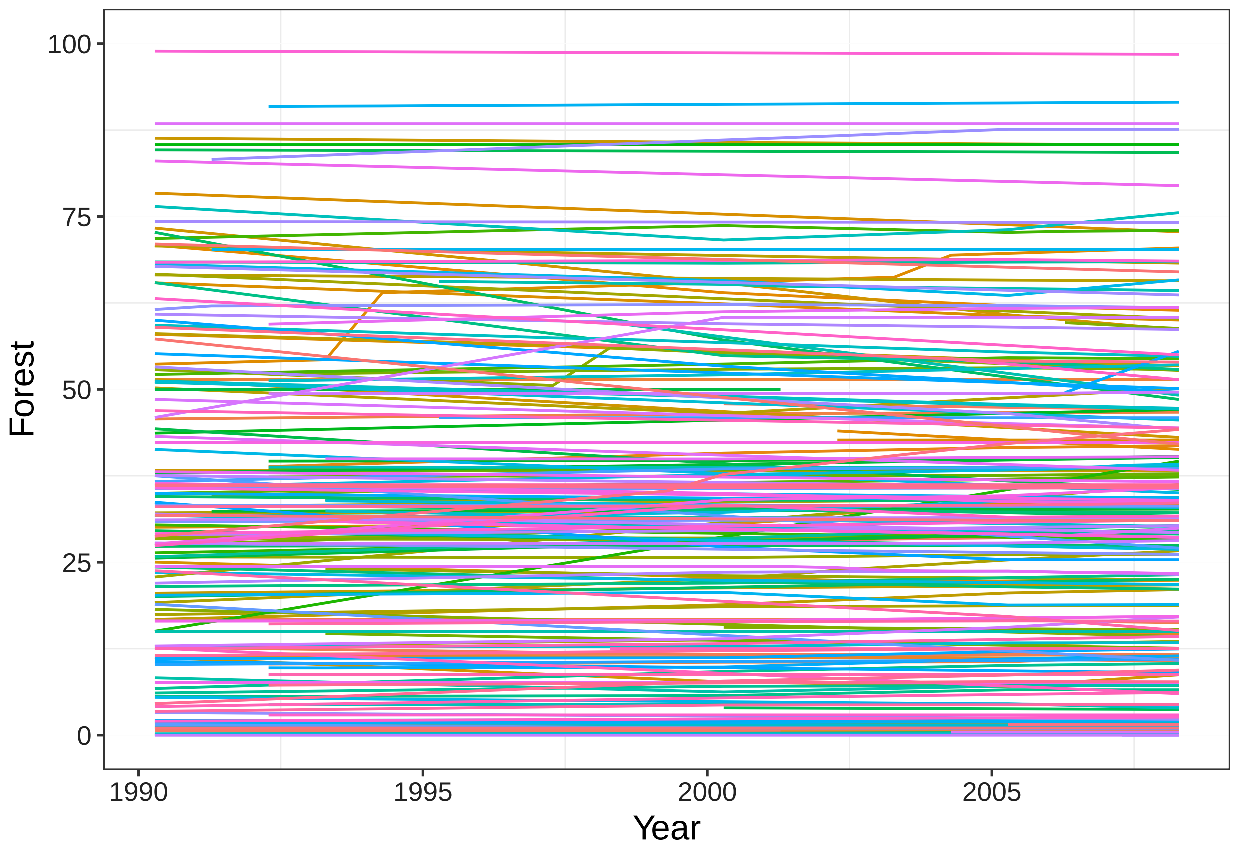
```r
summary(WB_Spread$Ag.Methane)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    1258    6596  155800   53060 3464398
```

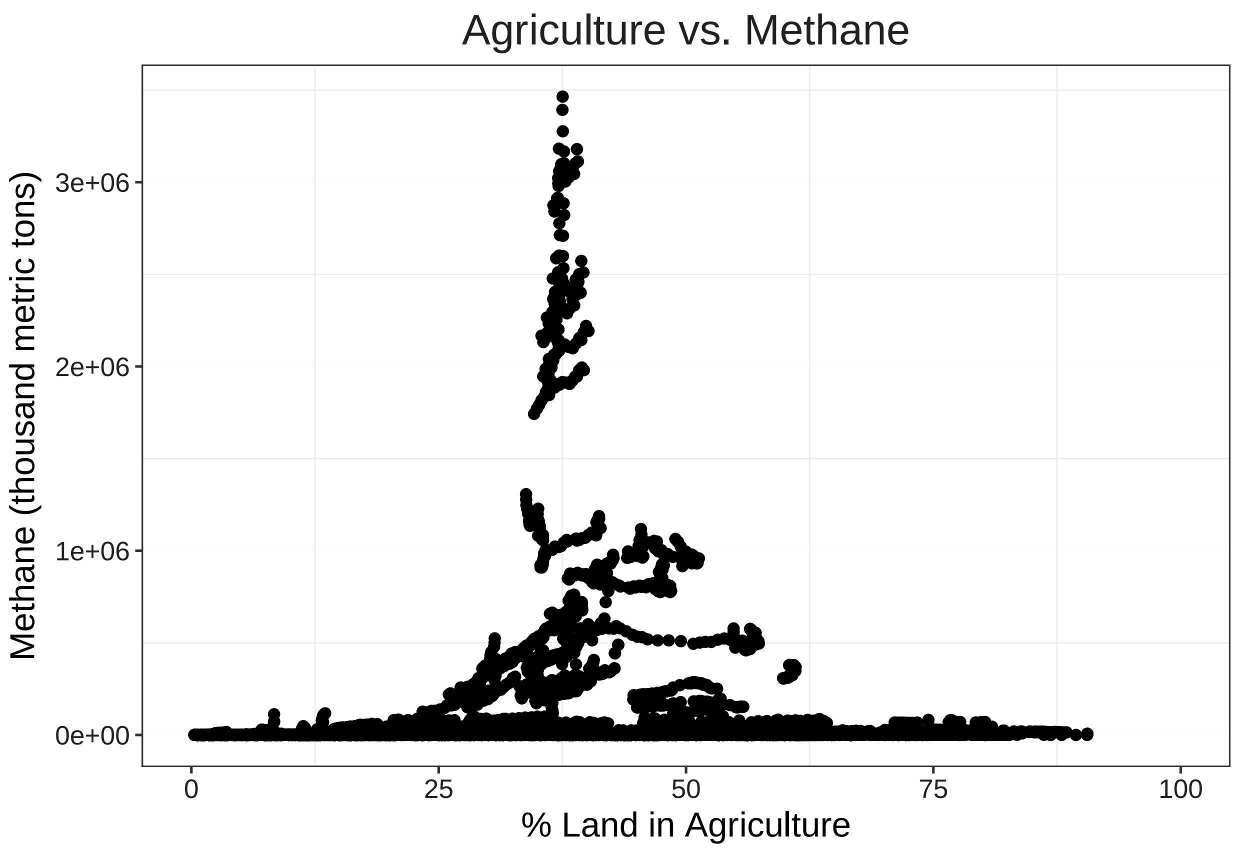
```
## Warning: Removed 8897 rows containing missing values (geom_point).
```



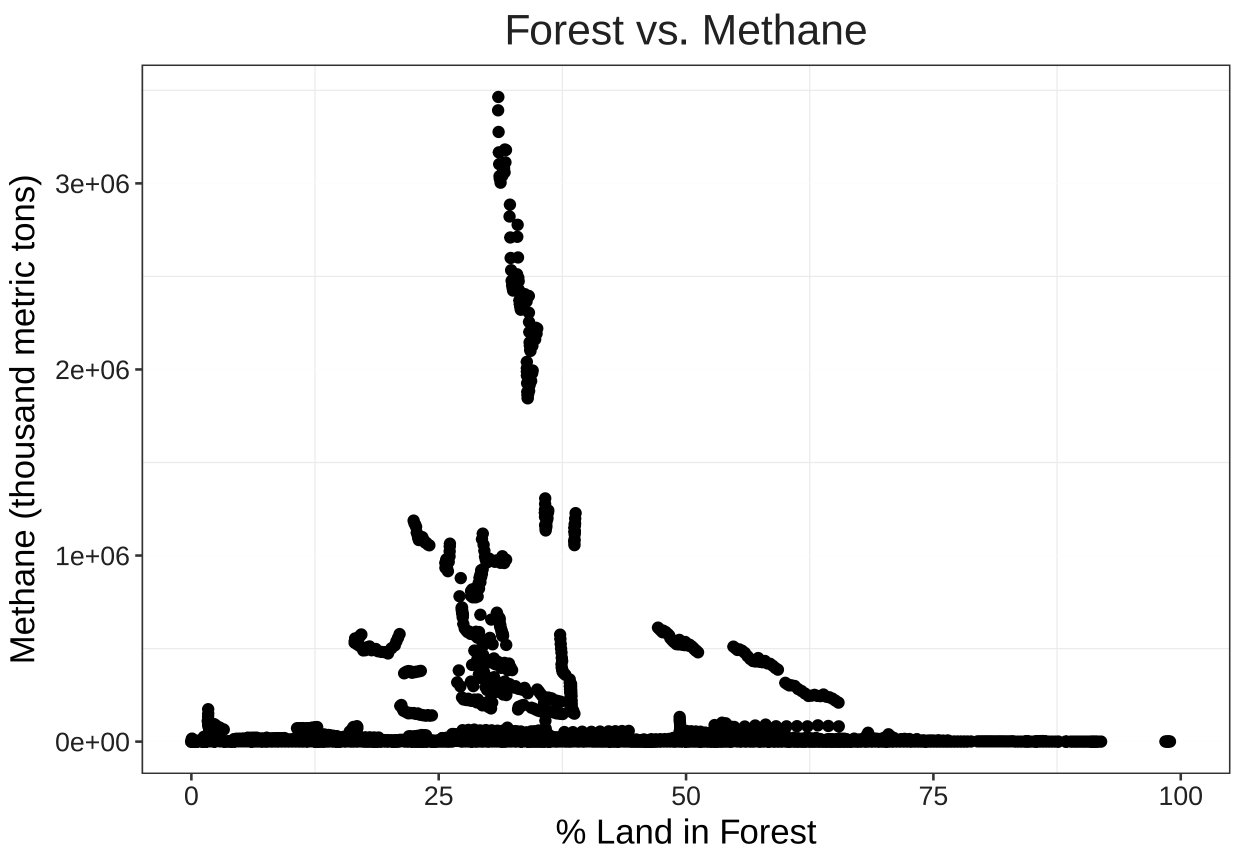

```
## Warning: Removed 33 rows containing missing values (geom_path).
```



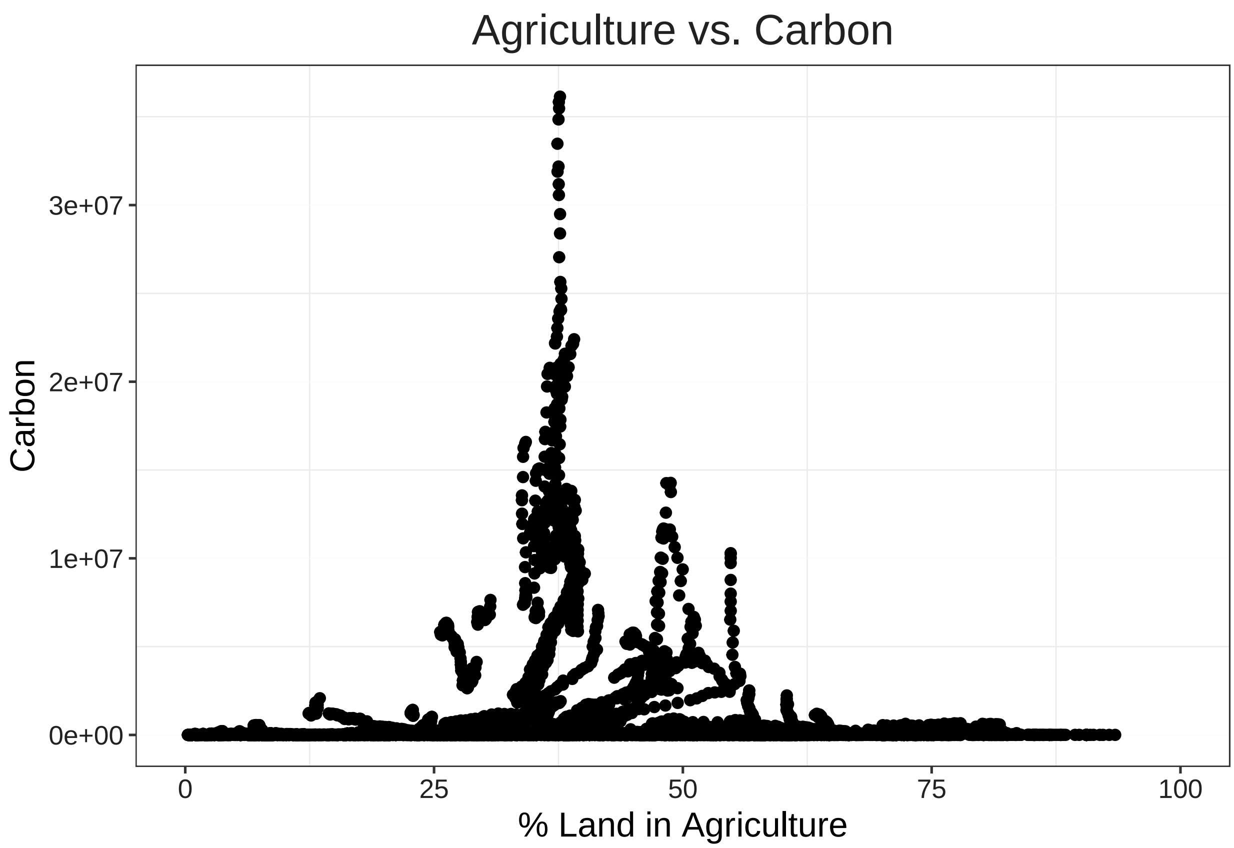

```
## Warning: Removed 6297 rows containing missing values (geom_point).
```



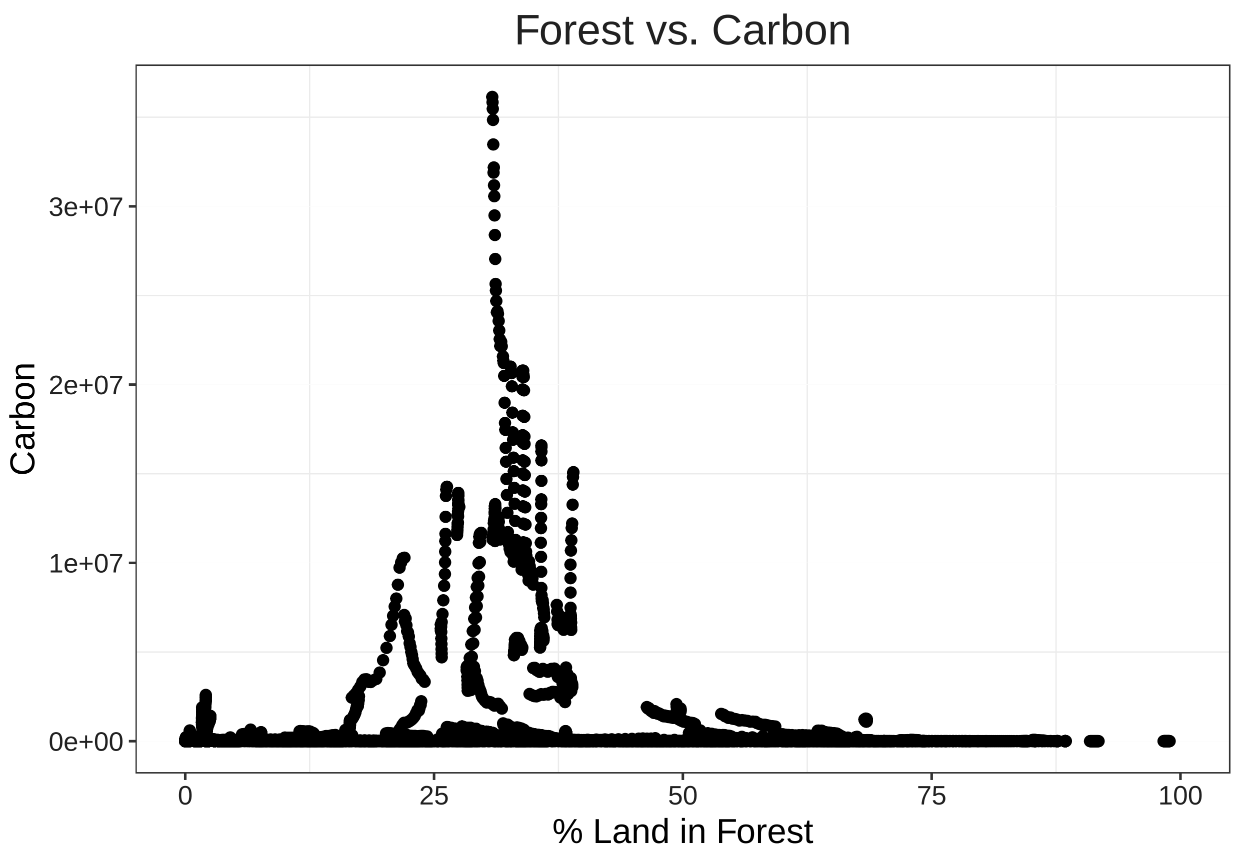

```
## Warning: Removed 10791 rows containing missing values (geom_point).
```




```
## Warning: Removed 3900 rows containing missing values (geom_point).
```

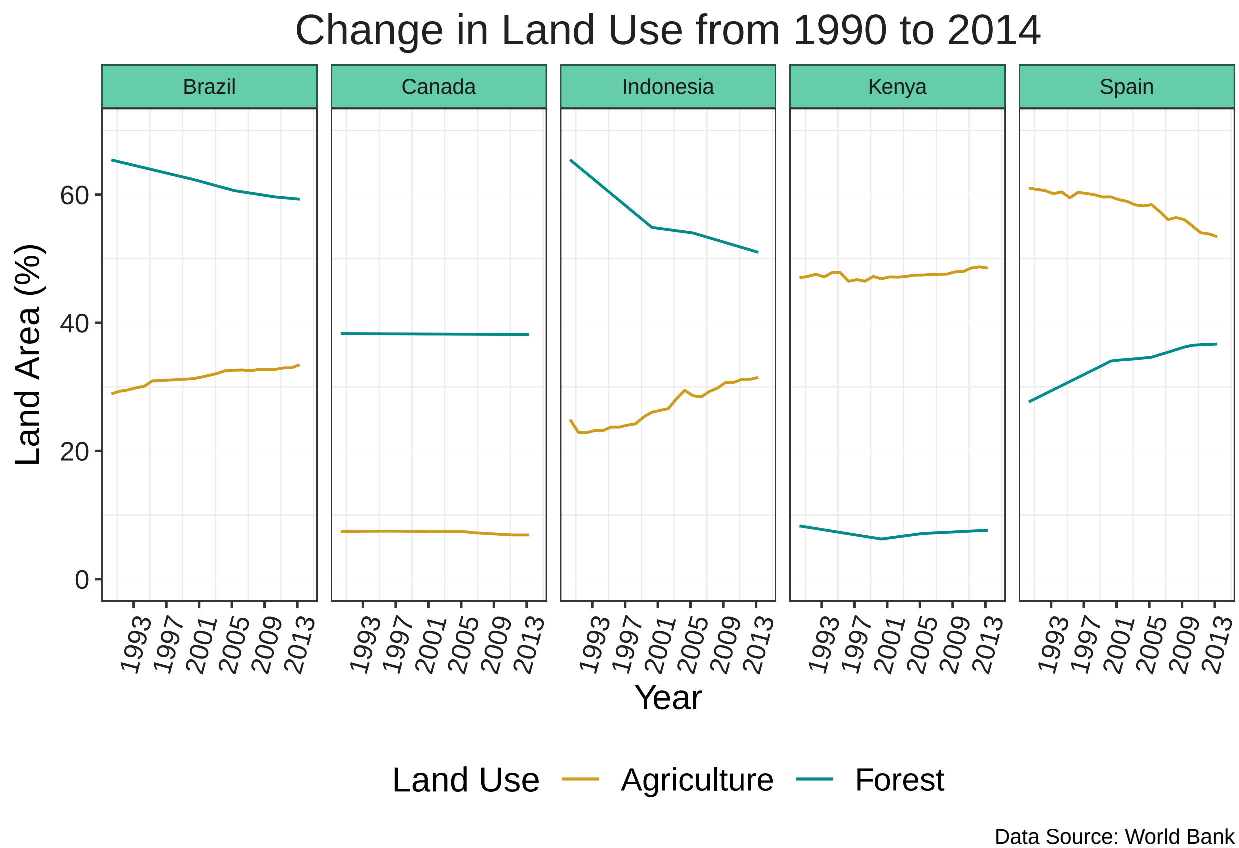



```
## Warning: Removed 9616 rows containing missing values (geom_point).
```



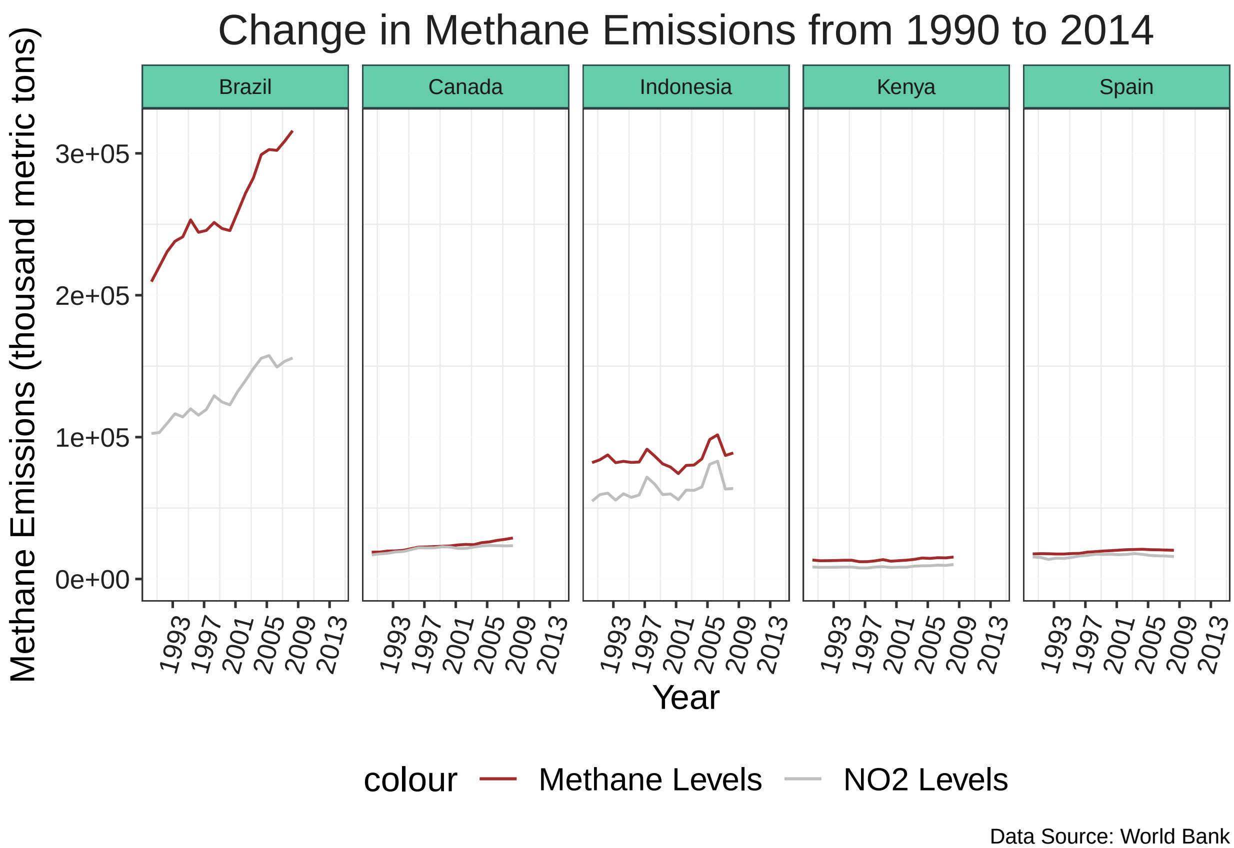

```
## Warning: Removed 35 rows containing missing values (geom_path).

## Warning: Removed 35 rows containing missing values (geom_path).
```

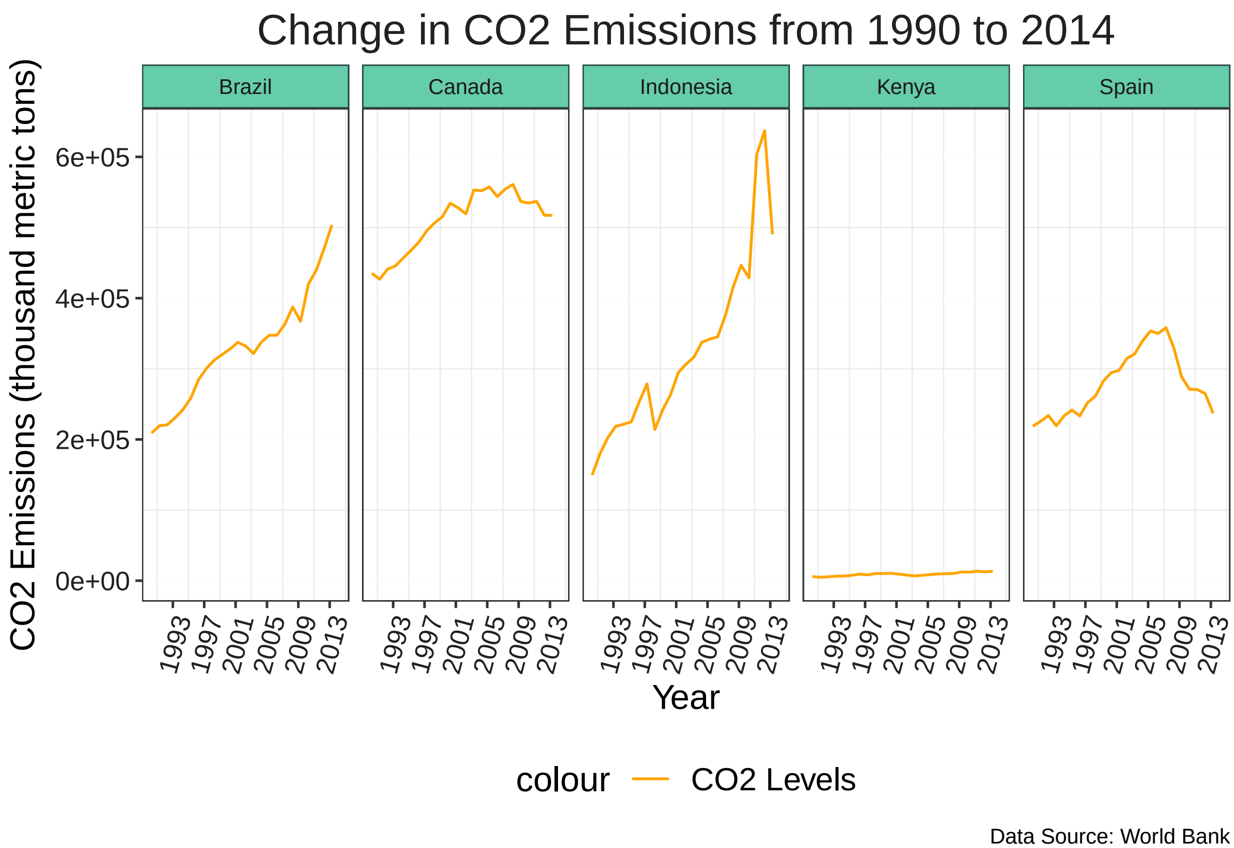



```
## Warning: Removed 40 rows containing missing values (geom_path).

## Warning: Removed 40 rows containing missing values (geom_path).
```

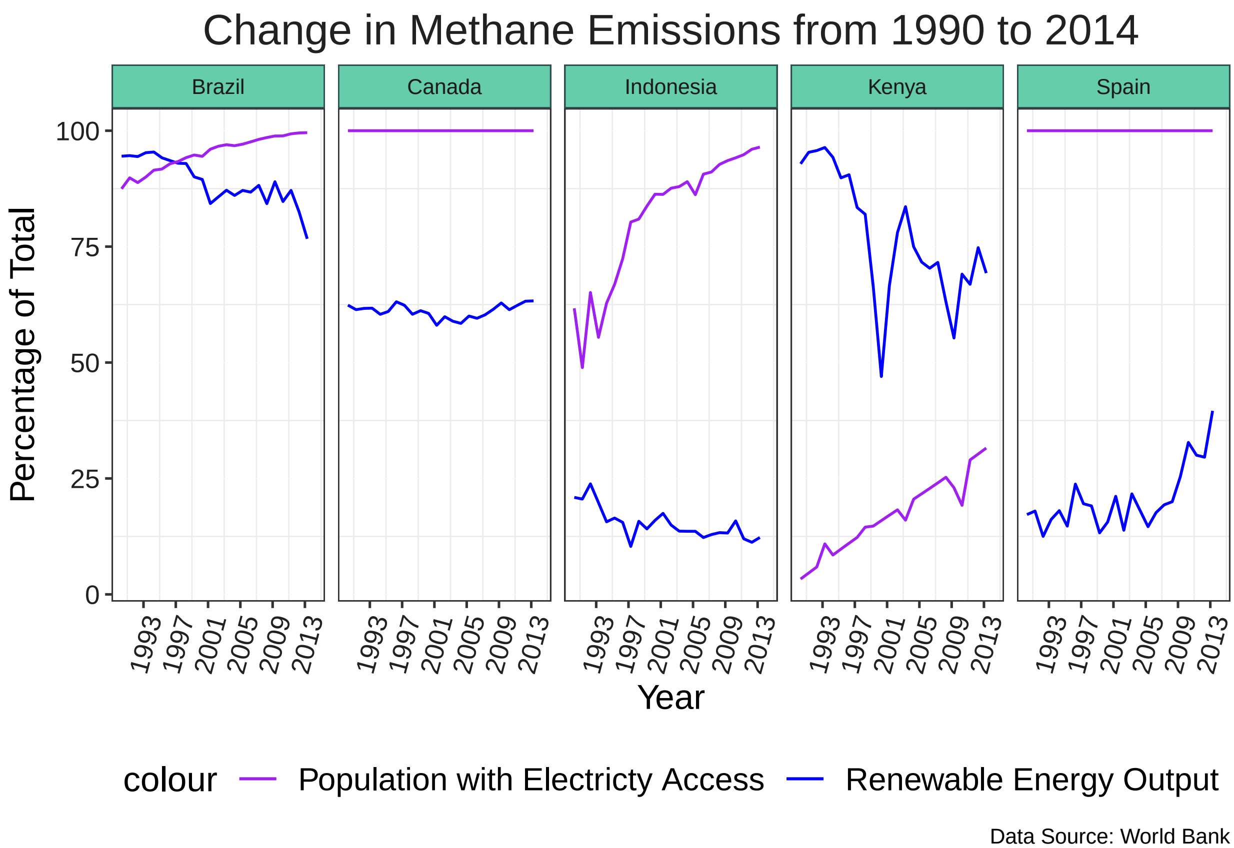



```
## Warning: Removed 35 rows containing missing values (geom_path).
```




```
## Warning: Removed 35 rows containing missing values (geom_path).

## Warning: Removed 35 rows containing missing values (geom_path).
```


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

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

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year
##   Data: WB_Spread
##        AIC     BIC    logLik
##   44741.09 44759.9 -22367.55
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 91.42949  6.525284 14.011572       0
## Year        -0.00485  0.000594 -8.156025       0
##
##  Correlation:
##      (Intr)
## Year -0.984
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -0.7540709 -0.3205935 -0.1002986  0.1413443 15.6053463
##
## Residual standard error: 73.63104
## Degrees of freedom: 3911 total; 3909 residual
```

```
#Did not do "* Country)". Result shows on average Forest decreases by -.00485 each yea
```

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.

```
Test1 <-  lme(data = WB_Spread,
              Forest ~ Year,
              random = ~1| Country)
Test1
```

```
## Linear mixed-effects model fit by REML
##   Data: WB_Spread
##   Log-restricted-likelihood: -22149.85
##   Fixed: Forest ~ Year
##  (Intercept)         Year
```

```
## 90.835100087 -0.004799758
##
## Random effects:
##  Formula: ~1 | Country
##         (Intercept) Residual
## StdDev:    30.78642 66.67633
##
## Number of Observations: 3911
## Number of Groups: 223
```

```r
#On average forest decreases by -.0047% a year?

Forest.Fixed_Ag. <- gls(data = WB_Spread,
                    Forest ~ Year + Agriculture,
                    method = "REML")
Forest.Fixed_Ag. #-.5067
```

```
## Generalized least squares fit by REML
##   Model: Forest ~ Year + Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -22331.49
##
## Coefficients:
##   (Intercept)          Year   Agriculture
## 110.607315028  -0.004830877  -0.506739618
##
## Degrees of freedom: 3911 total; 3908 residual
## Residual standard error: 72.92833
```

```r
#Random effect??
Test2 <-  lme(data = WB_Spread,
              Forest ~ Year + Agriculture,
              random = ~1| Country)
Test2 # -.557
```

```
## Linear mixed-effects model fit by REML
##   Data: WB_Spread
##   Log-restricted-likelihood: -22138.56
##   Fixed: Forest ~ Year + Agriculture
##   (Intercept)          Year   Agriculture
## 112.349788617  -0.004814667  -0.557561556
##
## Random effects:
##  Formula: ~1 | Country
##         (Intercept) Residual
## StdDev:    29.31384 66.60976
```

```
##
## Number of Observations: 3911
## Number of Groups: 223
```

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

```
## Warning in nlme::anova.lme(object = Forest.Fixed, Test2): fitted objects
## with different fixed effects. REML comparisons are not meaningful.
```

```
##              Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## Forest.Fixed     1  3 44741.09 44759.90 -22367.54
## Test2            2  5 44287.13 44318.48 -22138.56 1 vs 2 457.9608  <.0001
```

```
#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: -22287.92
##
## Coefficients:
##       (Intercept)              Year       Agriculture ElectricityAccess
##     129.772497937      -0.004179189      -0.560020533      -0.334328246
##
## Degrees of freedom: 3911 total; 3907 residual
## Residual standard error: 72.08392
```

```
#Added electricity access to see whether a lack of electricity may contribute to defor

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: -22357.42
##
## Coefficients:
##          (Intercept)                  Year RenewableElectricity
##         84.074720108          -0.004718623          0.181761849
##
## Degrees of freedom: 3911 total; 3908 residual
```

```
## Residual standard error: 73.40486
```

```r
#On that note, I was interested in whether there might be a relationship between fores

# 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
```

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

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

```
##
##  Pettitt's test for single change-point detection
##
## data:  WB_Spread$Forest
## U* = 458060, p-value = 1.459e-09
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                             1428
```

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

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
```

```r
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
```

Pettitt's applied to a single country (Brazil) initially detects a change point in Forest and Agriculture in the same year.

```r
#Statistical Test 3: Emmissions

AgMethane <- gls(data = WB_Spread,
                 Ag.Methane ~ Year + Agriculture,
                 method = "REML")
AgMethane #Ag. increases by 1, Ag Methane increases by 6.57
```

```
## Generalized least squares fit by REML
##   Model: Ag.Methane ~ Year + Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -56249.62
##
## Coefficients:
##   (Intercept)          Year   Agriculture
##  1.366512e+05 -5.517731e-01  6.577395e+02
##
## Degrees of freedom: 3911 total; 3908 residual
## Residual standard error: 428750.1
```

```r
ForestMethane <- gls(data = WB_Spread,
                  Ag.Methane ~ Year + Forest,
                  method = "REML")
ForestMethane #Forest increase by 1, methane increases by .035; Agricultural emmisions
```

```
## Generalized least squares fit by REML
##   Model: Ag.Methane ~ Year + Forest
##   Data: WB_Spread
##   Log-restricted-likelihood: -56245.46
##
## Coefficients:
```

```
##  (Intercept)         Year       Forest
## 1.289670e+05 1.195051e+00 3.563036e+02
##
## Degrees of freedom: 3911 total; 3908 residual
## Residual standard error: 428151.9
```

```r
Int.Methane <-  gls(data = WB_Spread,
                    Ag.Methane ~ Year + Forest * Agriculture,
                    method = "REML")
Int.Methane #Interaction of forest and agriculture: Increase of 1 leads to an increase
```

```
## Generalized least squares fit by REML
##   Model: Ag.Methane ~ Year + Forest * Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -56201.69
##
## Coefficients:
##        (Intercept)               Year             Forest
##       217760.79850            1.76024        -3223.93002
##        Agriculture Forest:Agriculture
##        -2149.63528           98.85697
##
## Degrees of freedom: 3911 total; 3906 residual
## Residual standard error: 424598.6
```

```r
Forest.CO2 <- gls(data = WB_Spread,
                  CO2Emissions ~ Year + Forest,
                  method = "REML")
Forest.CO2 #1% increase in forest leads to a -235.25 decrease in CO2
```

```
## Generalized least squares fit by REML
##   Model: CO2Emissions ~ Year + Forest
##   Data: WB_Spread
##   Log-restricted-likelihood: -63931.14
##
## Coefficients:
##  (Intercept)         Year       Forest
## 420051.95130     56.36262   -235.25547
##
## Degrees of freedom: 3911 total; 3908 residual
## Residual standard error: 3059879
```

```r
Ag.CO2 <- gls(data = WB_Spread,
              CO2Emissions ~ Year + Agriculture,
              method = "REML")
Ag.CO2 # 1% increase in Ag. leads to an increase of 1493 kt of carbon.
```

```
## Generalized least squares fit by REML
##   Model: CO2Emissions ~ Year + Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -63929.71
##
## Coefficients:
##  (Intercept)         Year  Agriculture
## 342001.55481     57.45718   1493.99738
##
## Degrees of freedom: 3911 total; 3908 residual
## Residual standard error: 3059780
```

```r
Int.CO2 <- gls(data = WB_Spread,
                    CO2Emissions ~ Year + Forest * Agriculture,
                    method = "REML")
Int.CO2 #How do you interperet this?
```

```
## Generalized least squares fit by REML
##   Model: CO2Emissions ~ Year + Forest * Agriculture
##   Data: WB_Spread
##   Log-restricted-likelihood: -63894.08
##
## Coefficients:
##        (Intercept)               Year             Forest
##      1130009.41628           59.26258       -22533.93863
##        Agriculture Forest:Agriculture
##       -17186.46246          611.63238
##
## Degrees of freedom: 3911 total; 3906 residual
## Residual standard error: 3042769
```

5 Summary and Conclusions