Every 15 Minutes

L Lathrop

1/19/2017

Introduction

In the summer of 2012, more than 100 maurauders on horseback crossed from Chad into Cameroon's Bouba Ndjidah National park with one mission: to kill as many elephants as possible without getting caught. Using rocket-propelled grenades and AK-47s, they mowed down hundreds of elephants, decimating entire herds, all for one reason: ivory.

The illegal killing of elephants in Africa for their ivory takes places sometimes on a small scale, with poisoned watermelons, and other times on a large scale, using belt-fed automatic weapons. But regardless of the method, when an elephant comes into contact with a poacher, more often than not, the poacher gets what he came for. This is why between 2010 and 2012 alone, some 100,000 elephants were slaughtered. There only "crime" that they hold the ivory that the world cannot seem to resist.

This project seeks to understand the socio-economic factors that may contribute to poaching. Several researchers have studied the problem, and their findings will be taken into account in the model. For instance, __________ studied the effect of regulated and unregulated markets and showed that the presence of unregulated markets, either in-country or in a bordering country, had a significant impact on poaching activity.

```r
library(ggplot2)
library(ggthemes)
library(psych)
library(GGally)
library(grDevices)
library(colorRamps)
library(dplyr)
library(maptools)
library(RColorBrewer)
library(blme)
library(lubridate)
library(reporttools)
library(stargazer)
library(fpp)
library(xtable)
library(plotly)
library(Cairo)
library(MASS)
library(car)
library(Amelia)
library(corrplot)
library(caret)
options(xtable.floating = FALSE)
options(xtable.timestamp = "")
```

Data preprocessing and exploratory analysis

```r
# READ IN DATASET
elephants <- read.csv("elephant_master.csv", header = TRUE)
```

```r
sapply(elephants, function(x) sum(is.na(x)))
```

```
##                      country                         year 
##                            0                            0 
##                         ISO2                         ISO3 
##                           14                            0 
##                       region                  subregionid 
##                            0                            0 
##                      cap.lat                     cap.long 
##                            0                            0 
##                    NGDP_RPCH                        NGDPD 
##                            2                            2 
##                      NGDPDPC                    NGSD_NGDP 
##                            2                           16 
##                         PCPI                      PCPIPCH 
##                            2                            2 
##                     GGX_NGDP                  GGXCNL_NGDP 
##                            2                            2 
##                  GGXWDG_NGDP                          BCA 
##                            2                            2 
##                          HDI                          GNI 
##                           34                           34 
##           Resource.Depletion               Adult.literacy 
##                          309                           70 
##       Primary.ed.enrollment               Mean.Schooling 
##                           98                          306 
##                    Total.pop         Pop.MultiDim.Povert 
##                           34                           71 
##       Deprivation.Intensity  Pop.Below.National.Poverty 
##                           71                           71 
##                  PPP.125.day        International.Dev.Aid 
##                           71                           34 
## Corruption.Perception.Index                   Reg.Market 
##                           34                            0 
##                 Unreg.Market         Reg.Market.Bordering 
##                            0                            0 
##      Unreg.Market.Bordering        Voice.Accountability 
##                            0                           34 
##          Political.Stability     Government.Effectiveness 
##                           34                           34 
##                     Rule.Law            Corruption.Control 
##                           34                           34 
##                  Reg.Quality               Armed.Conflict 
##                           34                            0 
##           Non.State.Conflict   Non.State.Conflict.Deaths 
##                            0                            0 
##                PIKE.regional            Definite.Possible 
##                           34                          374 
##               Elephant.range                Tot.Carcasses
```

```
##                          377                           174
##            Illegal.Carcasses
##                          174
```

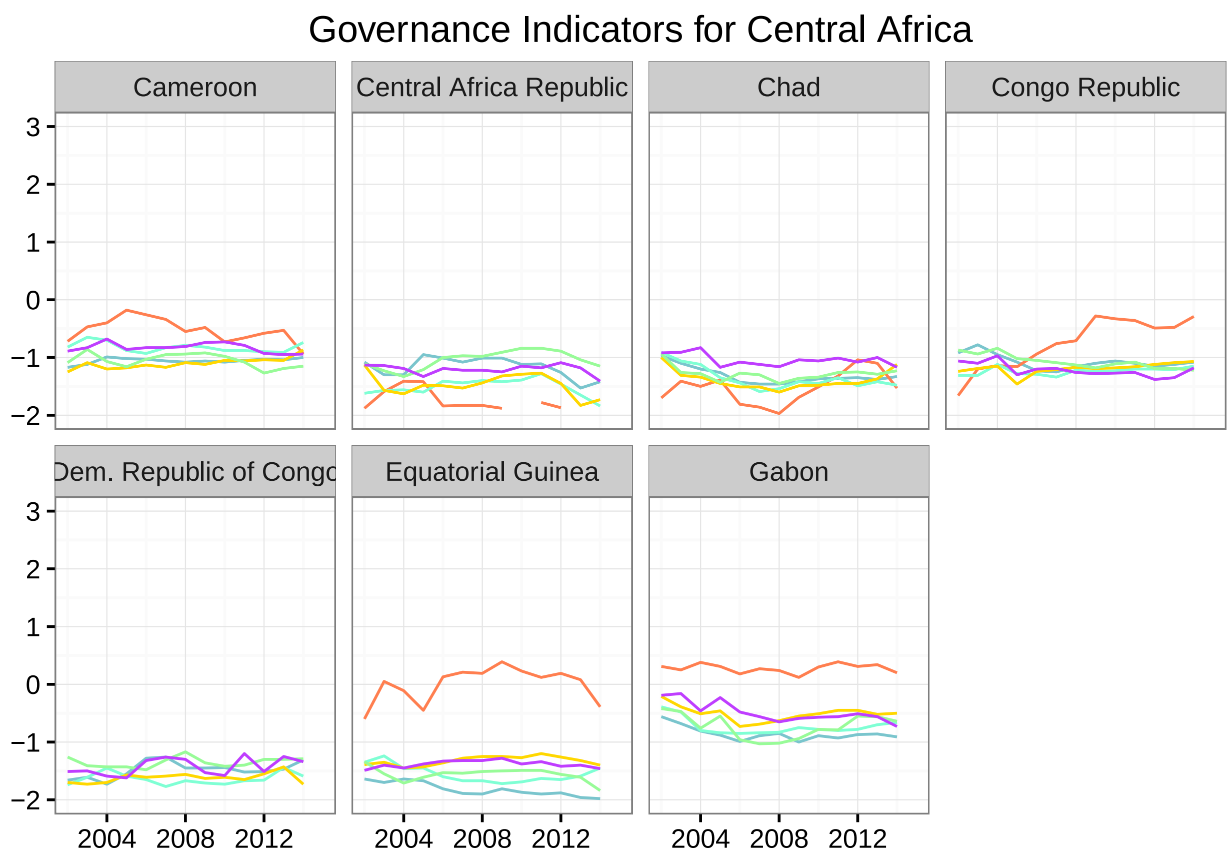
There are significant numbers of missing values in the dataset, so imputation will be necessary.

Note that the mean number of elephant carcasses for a given year is 49.1589404 and the mean number of illegal carcasses is 24.5066225, indicating that approximately half of all elephant deaths recorded are due to illegal poaching.

Governance indicators by region

```
elephantsFC <- filter(elephants, subregionid=="FC")
elephantsFW <- filter(elephants, subregionid=="FW")
elephantsFE <- filter(elephants, subregionid=="FE")
elephantsFS <- filter(elephants, subregionid=="FS")

governanceFC <- ggplot(data=elephantsFC, aes(x=year))+
  geom_line(aes(y=Voice.Accountability), color="cadetblue3")+
  geom_line(aes(y=Political.Stability), color="coral")+
  geom_line(aes(y=Government.Effectiveness), color="aquamarine")+
  geom_line(aes(y=Rule.Law), color="gold")+
  geom_line(aes(y=Corruption.Control), color="palegreen")+
  geom_line(aes(y=Reg.Quality), color="darkorchid1")+
  ylim(-2, 3) + xlab(NULL) + ylab(NULL) +
  ggtitle("Governance Indicators for Central Africa") + theme_bw()
governanceFC + facet_wrap(~country, ncol = 4)
```



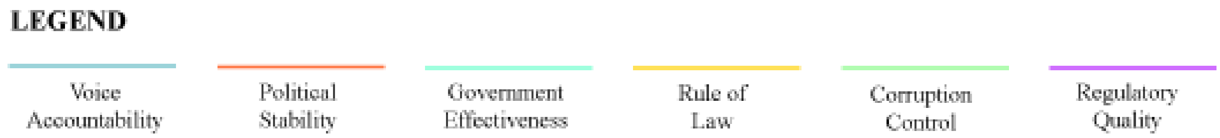
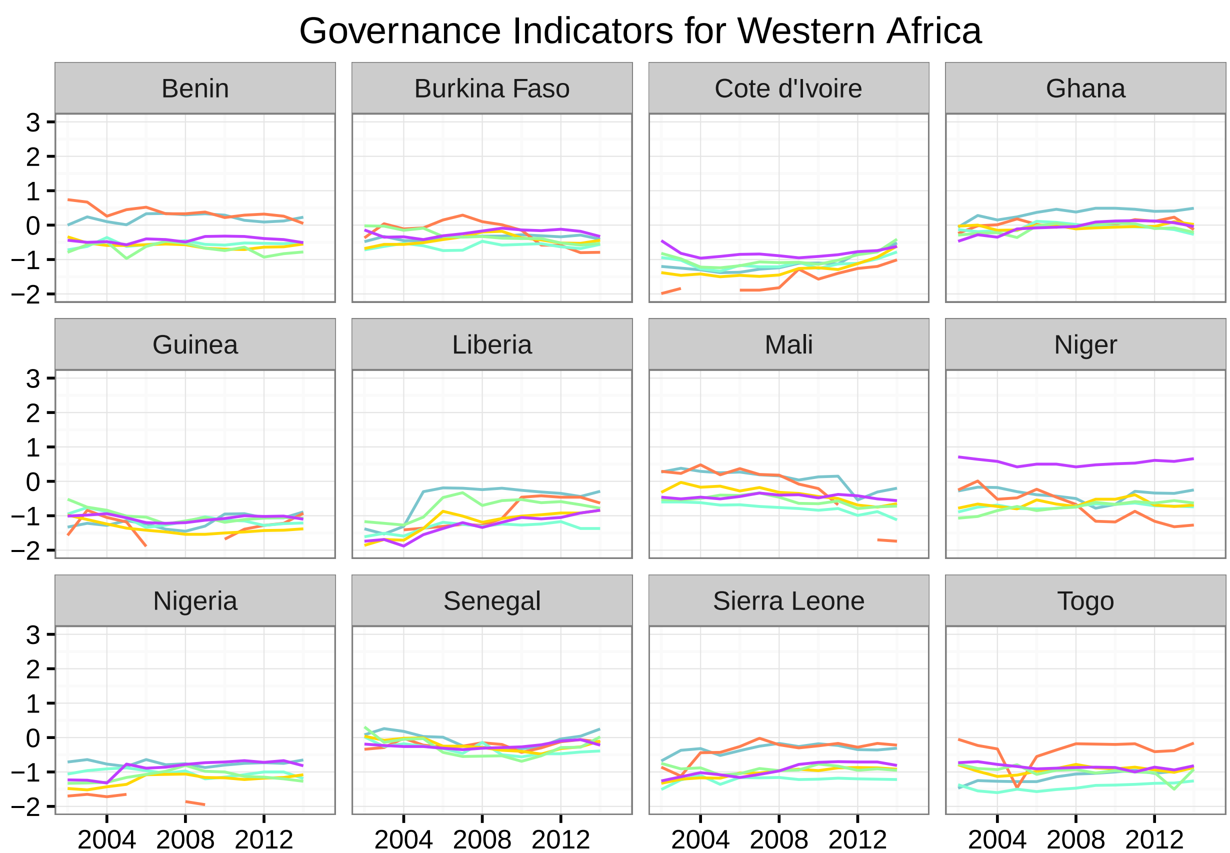


Figure 1:

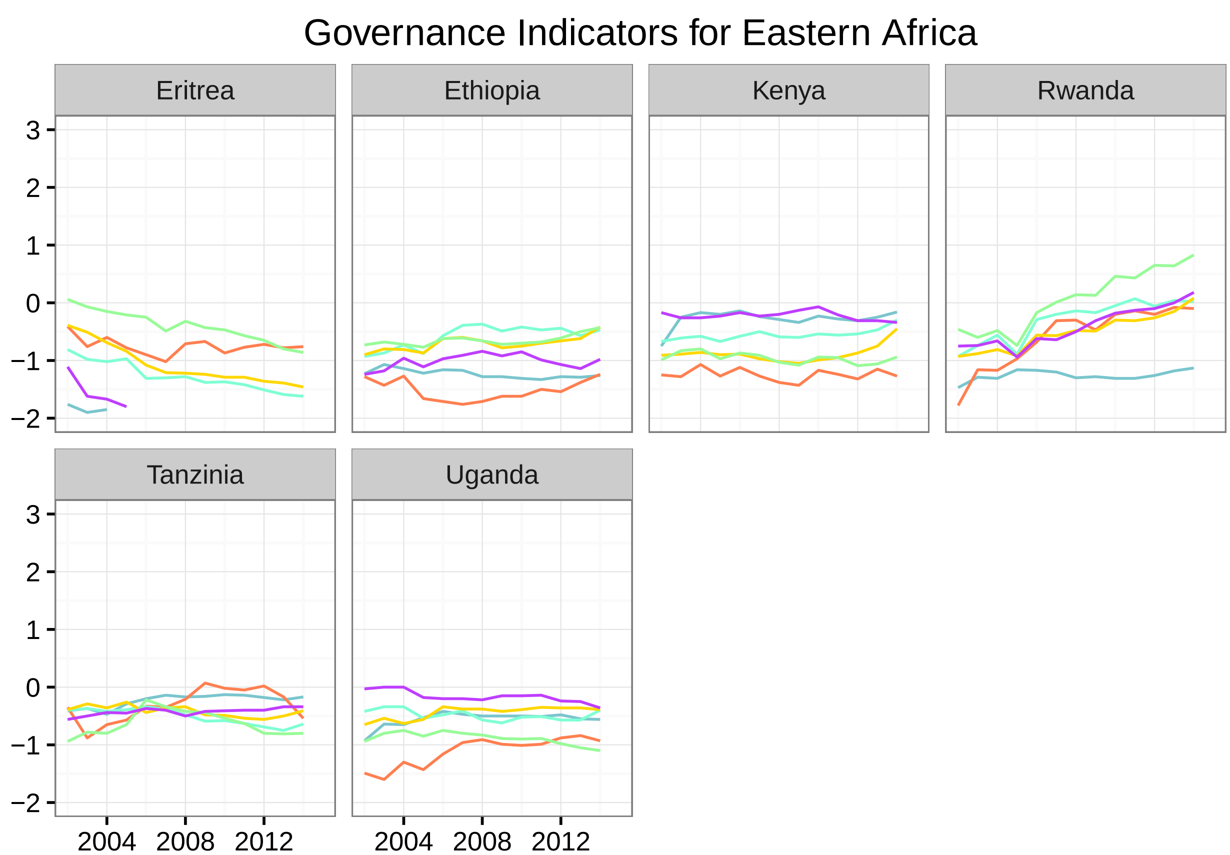
```
governanceFW <- ggplot(data=elephantsFW, aes(x=year))+
  geom_line(aes(y=Voice.Accountability), color="cadetblue3")+
  geom_line(aes(y=Political.Stability), color="coral")+
  geom_line(aes(y=Government.Effectiveness), color="aquamarine")+
  geom_line(aes(y=Rule.Law), color="gold")+
  geom_line(aes(y=Corruption.Control), color="palegreen")+
  geom_line(aes(y=Reg.Quality), color="darkorchid1")+
  ylim(-2, 3) + xlab(NULL) + ylab(NULL) +
  ggtitle("Governance Indicators for Western Africa")+ theme_bw()
governanceFW + facet_wrap(~country, ncol = 4)
```




```
governanceFE <- ggplot(data=elephantsFE, aes(x=year))+
  geom_line(aes(y=Voice.Accountability), color="cadetblue3")+
  geom_line(aes(y=Political.Stability), color="coral")+
  geom_line(aes(y=Government.Effectiveness), color="aquamarine")+
```

Figure 2:

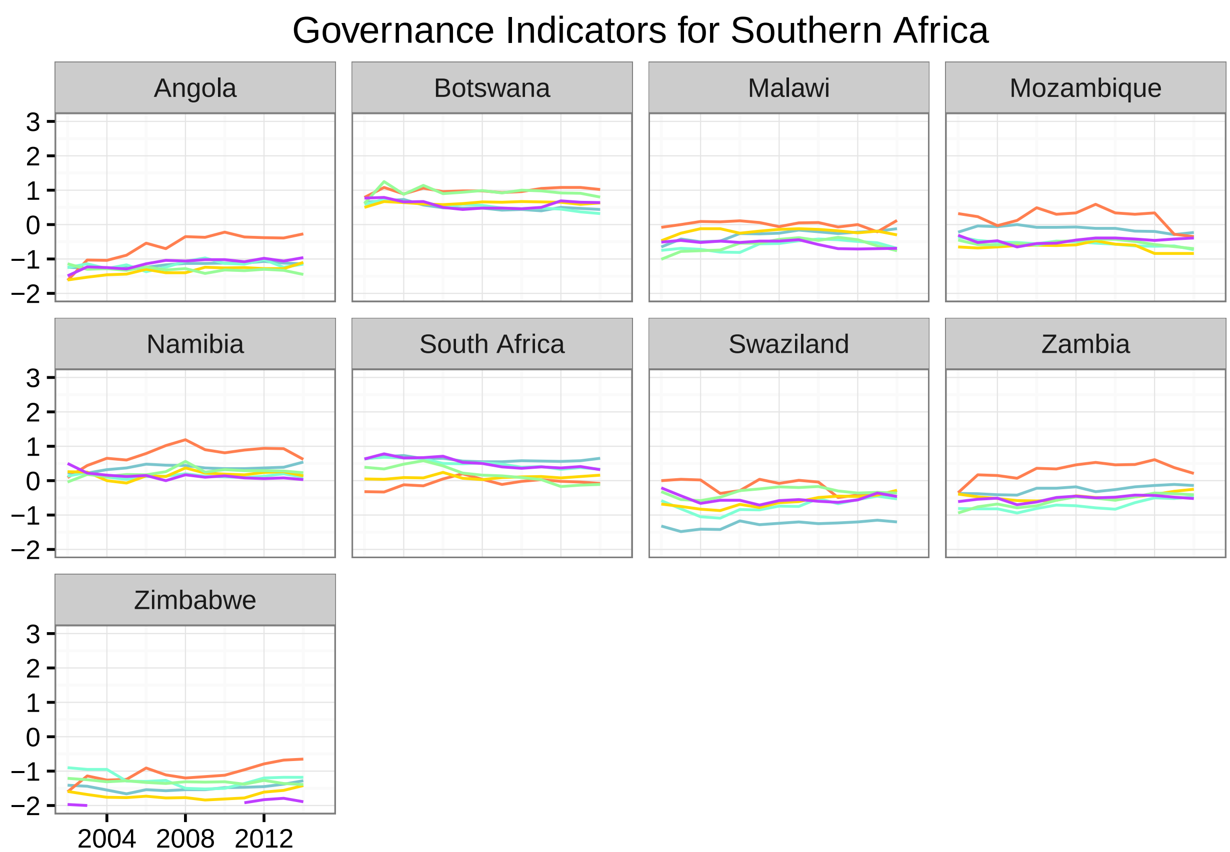
```r
  geom_line(aes(y=Rule.Law), color="gold")+
  geom_line(aes(y=Corruption.Control), color="palegreen")+
  geom_line(aes(y=Reg.Quality), color="darkorchid1")+
  ylim(-2, 3) + xlab(NULL) + ylab(NULL) +
  ggtitle("Governance Indicators for Eastern Africa")+ theme_bw()
governanceFE + facet_wrap(~country, ncol = 4)
```




```r
governanceFS <- ggplot(data=elephantsFS, aes(x=year))+
  geom_line(aes(y=Voice.Accountability), color="cadetblue3")+
  geom_line(aes(y=Political.Stability), color="coral")+
  geom_line(aes(y=Government.Effectiveness), color="aquamarine")+
  geom_line(aes(y=Rule.Law), color="gold")+
  geom_line(aes(y=Corruption.Control), color="palegreen")+
  geom_line(aes(y=Reg.Quality), color="darkorchid1")+
  ylim(-2, 3) + xlab(NULL) + ylab(NULL) +
```

Figure 3:

```
  ggtitle("Governance Indicators for Southern Africa")+ theme_bw()
governanceFS + facet_wrap(~country, ncol = 4)
```



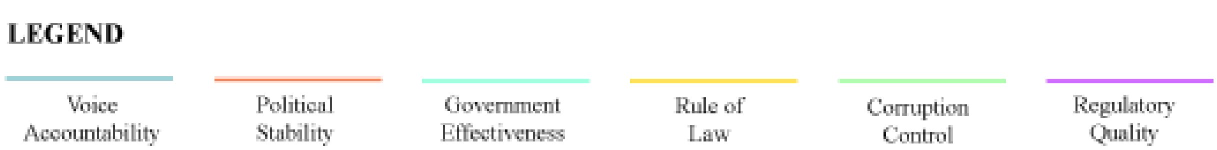


Figure 4:

From these visualizations we can see that governance indicators show a great deal of volatility, especially with respect to the countries of central Africa.

```r
# NORMALIZE AND STANDARDIZE THE DATA
centered.elephants <- data.frame(scale(elephants[,c(9:31,36:41,45:49)]))
centered.elephants <- cbind(elephants[,c(1:8, 32:35,42:44)], centered.elephants)
summary(centered.elephants)
```

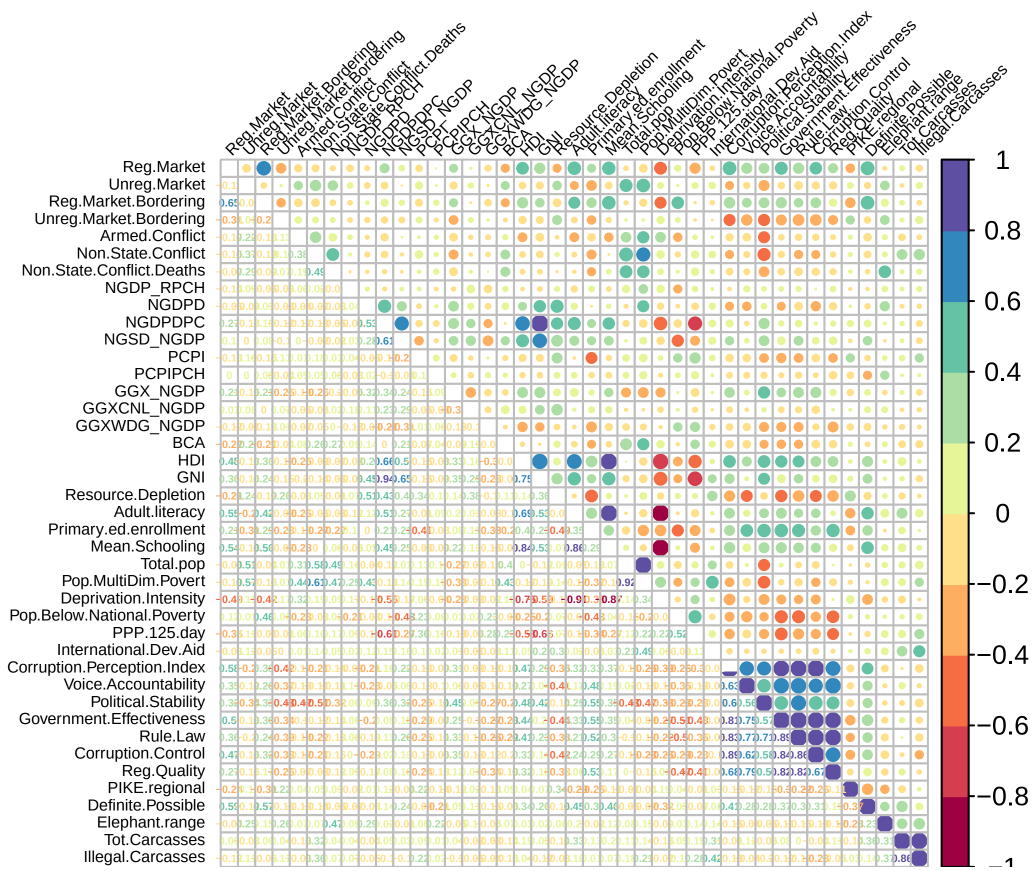
```
##                      country          year            ISO2           ISO3
##  Angola                  : 14   Min.   :2002   AO     : 14   AGO    : 14
##  Benin                   : 14   1st Qu.:2005   BF     : 14   BEN    : 14
##  Botswana                : 14   Median :2008   BJ     : 14   BFA    : 14
##  Burkina Faso            : 14   Mean   :2008   BW     : 14   BWA    : 14
##  Cameroon                : 14   3rd Qu.:2012   CD     : 14   CAF    : 14
##  Central Africa Republic: 14   Max.   :2015   (Other):392   CIV    : 14
##  (Other)                 :392                 NA's   : 14   (Other):392
##             region    subregionid    cap.lat            cap.long
##  Central Africa : 98   FC: 98      Min.   :-26.180   Min.   :-17.29
##  Eastern Africa : 84   FE: 84      1st Qu.: -8.500   1st Qu.:  1.20
##  Southern Africa:126   FS:126      Median :  3.475   Median : 14.86
##  Western Africa :168   FW:168      Mean   : -1.042   Mean   : 14.31
##                                    3rd Qu.:  9.020   3rd Qu.: 31.02
##                                    Max.   : 15.190   Max.   : 38.55
##
##    Reg.Market      Unreg.Market    Reg.Market.Bordering
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.1176   Mean   :0.2059   Mean   :0.4706
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :3.0000
##
##  Unreg.Market.Bordering Armed.Conflict   Non.State.Conflict
##  Min.   :0.000          Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.000          1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.000          Median :0.0000   Median :0.0000
##  Mean   :1.412          Mean   :0.1744   Mean   :0.1429
##  3rd Qu.:2.000          3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :4.000          Max.   :1.0000   Max.   :1.0000
##
##  Non.State.Conflict.Deaths   NGDP_RPCH           NGDPD
##  Min.   :   0.00           Min.   :-7.6486   Min.   :-0.2287
##  1st Qu.:   0.00           1st Qu.:-0.3971   1st Qu.:-0.2259
##  Median :   0.00           Median : 0.0240   Median :-0.2220
##  Mean   :  41.52           Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.:   0.00           3rd Qu.: 0.3883   3rd Qu.:-0.2133
##  Max.   :2969.00           Max.   : 6.1413   Max.   : 7.7599
##                            NA's   :2         NA's   :2
##     NGDPDPC           NGSD_NGDP             PCPI             PCPIPCH
##  Min.   :-0.56879   Min.   :-5.08625   Min.   :-0.5900   Min.   :-6.0292
##  1st Qu.:-0.47391   1st Qu.:-0.55617   1st Qu.:-0.3679   1st Qu.:-0.4219
##  Median :-0.38399   Median :-0.06108   Median :-0.2965   Median :-0.1483
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.:-0.08128   3rd Qu.: 0.41464   3rd Qu.:-0.1159   3rd Qu.: 0.2187
##  Max.   : 6.74896   Max.   : 4.61274   Max.   : 6.6623   Max.   :11.1132
##  NA's   :2          NA's   :16         NA's   :2         NA's   :2
##     GGX_NGDP         GGXCNL_NGDP       GGXWDG_NGDP            BCA
```

```
##  Min.   :-2.2867   Min.   :-3.7948   Min.   :-0.71674   Min.   :-4.45160
##  1st Qu.:-0.6501   1st Qu.:-0.4466   1st Qu.:-0.41669   1st Qu.:-0.17566
##  Median :-0.1903   Median :-0.1239   Median :-0.24927   Median :-0.01854
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.4814   3rd Qu.: 0.2182   3rd Qu.: 0.04561   3rd Qu.: 0.04971
##  Max.   : 4.7642   Max.   : 6.4128   Max.   : 9.47498   Max.   : 7.87090
##  NA's   :2         NA's   :2         NA's   :2          NA's   :2
##       HDI               GNI           Resource.Depletion Adult.literacy
##  Min.   :-1.9747   Min.   :-0.69636   Min.   :-0.79871   Min.   :-1.7751
##  1st Qu.:-0.7177   1st Qu.:-0.53992   1st Qu.:-0.61896   1st Qu.:-0.6691
##  Median :-0.1539   Median :-0.38762   Median :-0.34798   Median : 0.2316
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.5703   3rd Qu.:-0.00254   3rd Qu.: 0.05983   3rd Qu.: 0.5686
##  Max.   : 2.5877   Max.   : 5.89860   Max.   : 3.51010   Max.   : 1.5693
##  NA's   :34        NA's   :34         NA's   :309        NA's   :70
##  Primary.ed.enrollment Mean.Schooling      Total.pop
##  Min.   :-2.3427       Min.   :-1.5998   Min.   :-0.7197
##  1st Qu.:-0.7217       1st Qu.:-0.7544   1st Qu.:-0.5807
##  Median : 0.4164       Median :-0.1638   Median :-0.3005
##  Mean   : 0.0000       Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.8494       3rd Qu.: 0.6237   3rd Qu.: 0.0131
##  Max.   : 1.2546       Max.   : 2.3840   Max.   : 5.2995
##  NA's   :98            NA's   :306       NA's   :34
##  Pop.MultiDim.Povert Deprivation.Intensity Pop.Below.National.Poverty
##  Min.   :-0.7126     Min.   :-2.0053       Min.   :-2.11053
##  1st Qu.:-0.5264     1st Qu.:-0.5484       1st Qu.:-0.54306
##  Median :-0.3401     Median :-0.1437       Median :-0.02057
##  Mean   : 0.0000     Mean   : 0.0000       Mean   : 0.00000
##  3rd Qu.:-0.1183     3rd Qu.: 0.5847       3rd Qu.: 0.63255
##  Max.   : 3.4444     Max.   : 2.2683       Max.   : 1.94646
##  NA's   :71          NA's   :71            NA's   :71
##   PPP.125.day      International.Dev.Aid Corruption.Perception.Index
##  Min.   :-2.0776   Min.   :-0.35140      Min.   :-1.5699
##  1st Qu.:-0.5815   1st Qu.:-0.28960      1st Qu.:-0.6889
##  Median :-0.1415   Median :-0.17179      Median :-0.2484
##  Mean   : 0.0000   Mean   : 0.00000      Mean   : 0.0000
##  3rd Qu.: 0.7489   3rd Qu.: 0.06152      3rd Qu.: 0.4123
##  Max.   : 2.1467   Max.   :18.50638      Max.   : 3.9364
##  NA's   :71        NA's   :34            NA's   :34
##  Voice.Accountability Political.Stability Government.Effectiveness
##  Min.   :-2.16869     Min.   :-2.4849     Min.   :-1.89661
##  1st Qu.:-0.82092     1st Qu.:-0.7909     1st Qu.:-0.79337
##  Median :-0.03951     Median : 0.2065     Median :-0.00534
##  Mean   : 0.00000     Mean   : 0.0000     Mean   : 0.00000
##  3rd Qu.: 0.69889     3rd Qu.: 0.7424     3rd Qu.: 0.51701
##  Max.   : 2.00364     Max.   : 2.1119     Max.   : 2.73250
##  NA's   :34           NA's   :34          NA's   :34
##     Rule.Law        Corruption.Control  Reg.Quality
##  Min.   :-1.91831   Min.   :-2.0943    Min.   :-2.50144
##  1st Qu.:-0.83073   1st Qu.:-0.6954    1st Qu.:-0.65564
##  Median : 0.02881   Median :-0.1595    Median : 0.04045
##  Mean   : 0.00000   Mean   : 0.0000    Mean   : 0.00000
##  3rd Qu.: 0.66031   3rd Qu.: 0.4673    3rd Qu.: 0.59184
##  Max.   : 2.51972   Max.   : 3.5194    Max.   : 2.26949
```

```
##  NA's   :34         NA's   :34         NA's   :34
##  PIKE.regional     Definite.Possible Elephant.range    Tot.Carcasses
##  Min.   :-1.5287   Min.   :-0.4492   Min.   :-0.7018   Min.   :-0.52983
##  1st Qu.:-0.7476   1st Qu.:-0.4461   1st Qu.:-0.6203   1st Qu.:-0.48672
##  Median :-0.2904   Median :-0.3970   Median :-0.4936   Median :-0.37355
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 1.0136   3rd Qu.:-0.0240   3rd Qu.: 0.2595   3rd Qu.:-0.03405
##  Max.   : 1.7035   Max.   : 4.5039   Max.   : 5.2583   Max.   : 7.09011
##  NA's   :34        NA's   :374       NA's   :377       NA's   :174
##  Illegal.Carcasses
##  Min.   :-0.4849
##  1st Qu.:-0.4651
##  Median :-0.3860
##  Mean   : 0.0000
##  3rd Qu.:-0.0496
##  Max.   : 8.3991
##  NA's   :174
```

Several of the variables still appear to be highly skewed, so further transformation will be required, as the `Amelia II` package imputes missing values assuming a normal distribution.

```
# EXAMINE DATA FOR HIGHLY CORRELATED VARIABLES
clrs <- brewer.pal(10, "Spectral")
cors <- cor(centered.elephants[,9:49], use = "pairwise")
corrplot <- corrplot.mixed(cors, col = clrs, number.cex=0.3,
                    tl.pos="lt", tl.cex=0.5, tl.col="black",
                    tl.srt=45)
```



```r
# REMOVE HIGHLY CORRELATED VARIABLES
pd.elephants <- subset(centered.elephants,
                       subset=TRUE, select=c(1:10,12:23,26:40,42:49))

# CREATE DUMMY VARIABLE TO USE FOR BINOMIAL RESPONSE
pd.elephants$Percent.Illegal <- pd.elephants$Illegal.Carcasses /
                                  pd.elephants$Tot.Carcasses

pd.elephants$High.Illegal <- factor(with(pd.elephants,
                                         ifelse((Percent.Illegal > 0.5), 1, 0)))

table(pd.elephants$High.Illegal)
```

```
##
##   0   1
##  50 252
```

Linear regression was run on each of the highly correlated variables to determine which variables should be removed. Four variables were deleted from the dataset. We also see that of the observed values of total elephant carasses in relation to illegal carcasses, 252 observations meet the condition of more than 50% of the total were killed illegally.

```r
# SEPARATE DATA INTO TRAINING AND TESTING SETS
inVal <- createDataPartition(pd.elephants$High.Illegal, p = 0.7, list=FALSE)
eleph.train <- pd.elephants[inVal,]
eleph.test <- pd.elephants[-inVal,]

# IMPUTE MISSING VALUES
library(snow)
set.seed(1357)
elephants.out.train <- amelia(eleph.train, m=5, frontend = FALSE,
                              idvars = c("ISO2", "ISO3", "region",
                                         "subregionid", "cap.lat", "cap.long"),
                              ts = "year", cs = "country", noms = "High.Illegal",
                              logs = c("Tot.Carcasses", "Illegal.Carcasses"),
                              polytime = 0, intercs = TRUE, p2s = 1,
                              parallel="snow", ncpus = 3,
                             empri = .01*nrow(centered.elephants))

# Amelia II: Multiple Imputation
# (Version 1.7.4, built: 2015-12-05)
# Copyright (C) 2005-2017 James Honaker, Gary King and Matthew Blackwell
# Refer to http://gking.harvard.edu/amelia/ for more information
```

```r
# SPOT CHECK VARIABLES WITH HIGH MISSINGNESS
par(mfrow=c(3,2))
tscsPlot(elephants.out.train, cs = "Swaziland", main = "Swaziland",
          var = "Resource.Depletion", ylim = c(-2,1))
```

```
## Warning in amcheck(x = x, m = m, idvars = numopts$idvars, priors = priors, : The log transformation i
##  variables with negative values. The values
##  will be shifted up by 1 plus the minimum value
##  of that variable.
```

```r
tscsPlot(elephants.out.train, cs = "Botswana", main = "Botswana",
         var = "Resource.Depletion", ylim = c(-2,1))
```

```
## Warning in amcheck(x = x, m = m, idvars = numopts$idvars, priors = priors, : The log transformation i
##  variables with negative values. The values
##  will be shifted up by 1 plus the minimum value
##  of that variable.
```

```r
tscsPlot(elephants.out.train, cs = "Cameroon", main = "Cameroon",
         var = "Definite.Possible", ylim = c(-2,1))
```

```
## Warning in amcheck(x = x, m = m, idvars = numopts$idvars, priors = priors, : The log transformation i
##  variables with negative values. The values
##  will be shifted up by 1 plus the minimum value
##  of that variable.
```

```r
tscsPlot(elephants.out.train, cs = "Central Africa Republic",
         main = "Central Africa Republic",
         var = "Definite.Possible", ylim = c(-2,1))
```

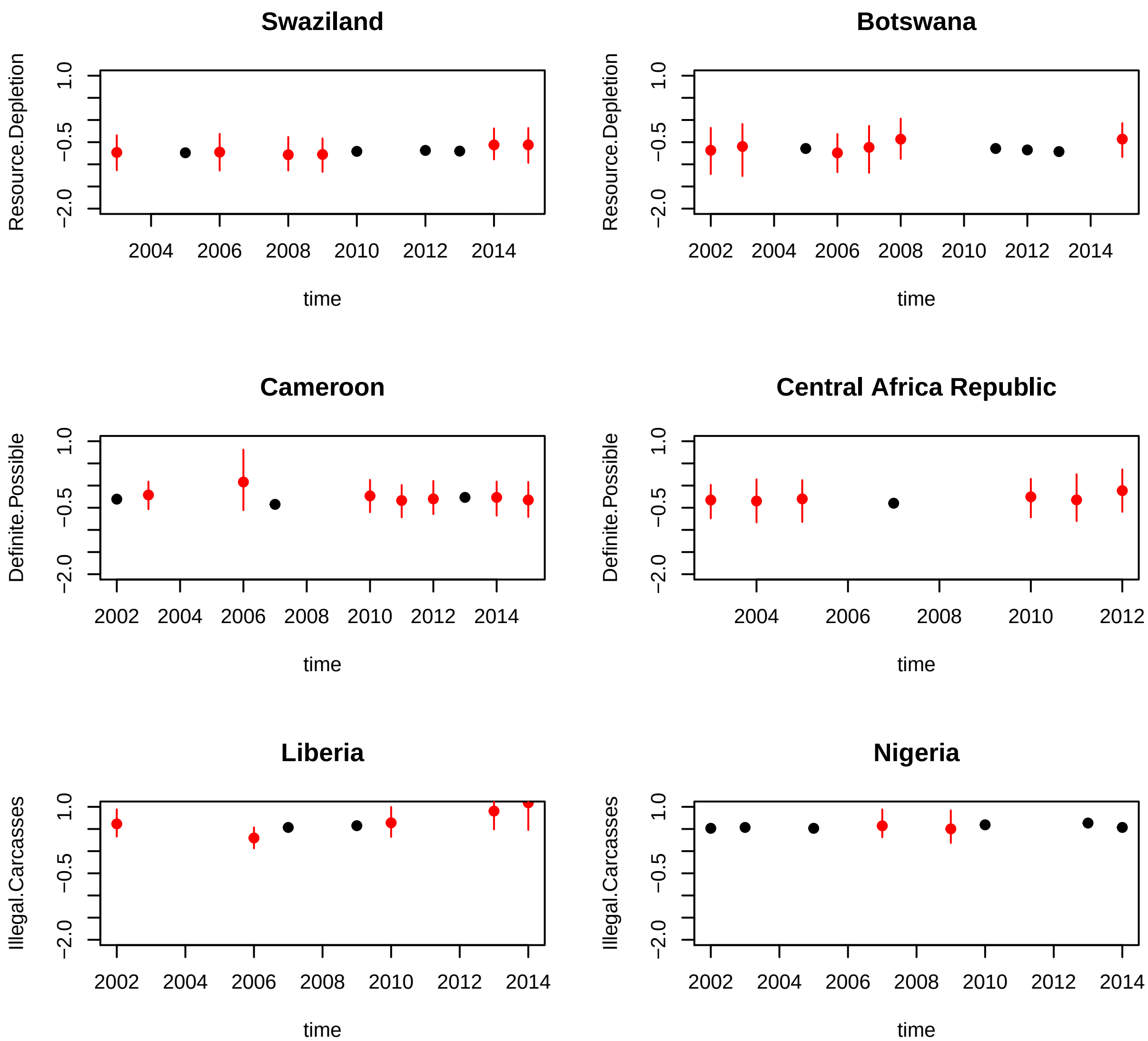
```
## Warning in amcheck(x = x, m = m, idvars = numopts$idvars, priors = priors, : The log transformation i
##  variables with negative values. The values
##  will be shifted up by 1 plus the minimum value
##  of that variable.
```

```r
tscsPlot(elephants.out.train, cs = "Liberia", main = "Liberia",
         var = "Illegal.Carcasses", ylim = c(-2,1))
```

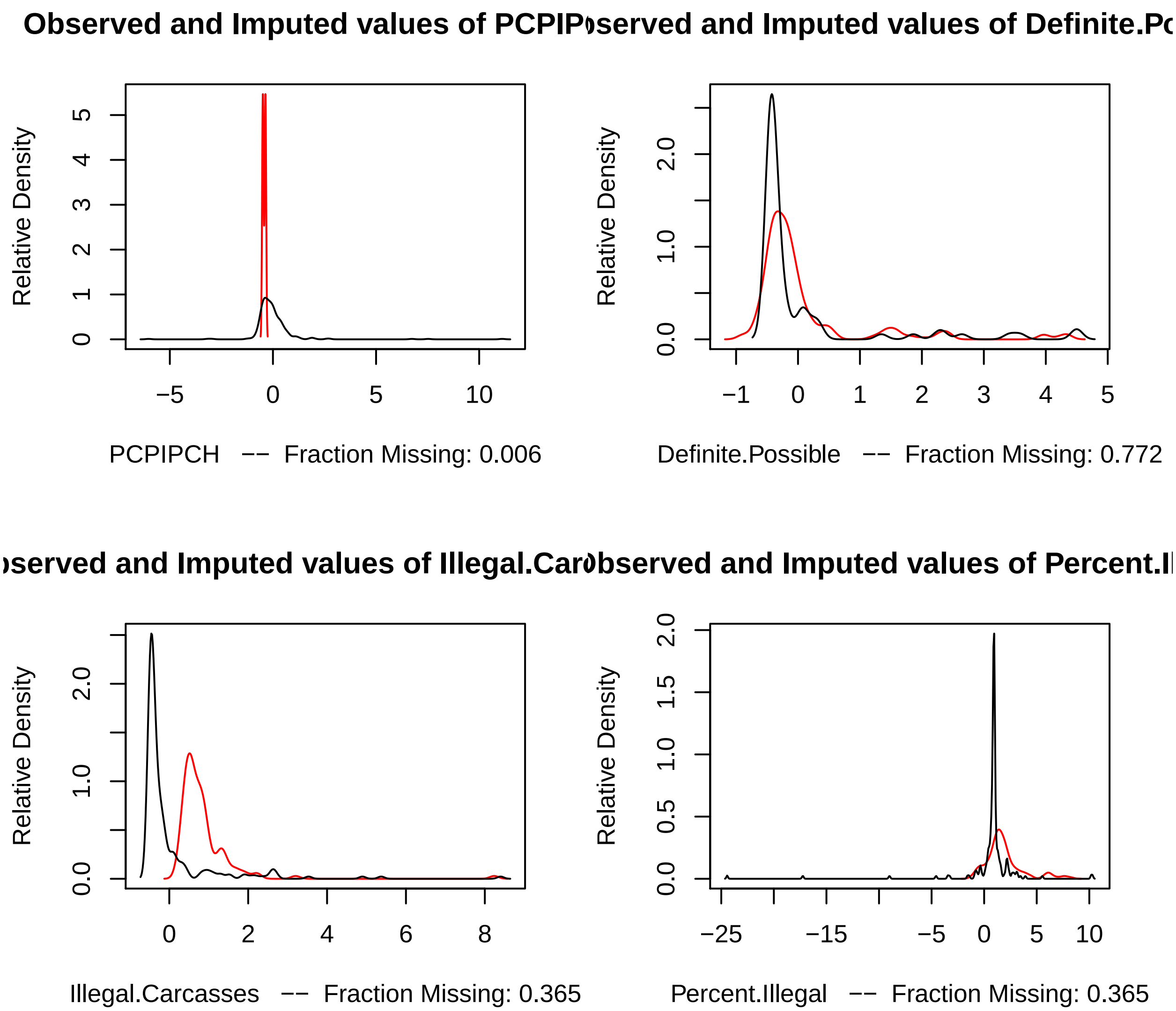
```
## Warning in amcheck(x = x, m = m, idvars = numopts$idvars, priors = priors, : The log transformation i
##  variables with negative values. The values
##  will be shifted up by 1 plus the minimum value
##  of that variable.
```

```r
tscsPlot(elephants.out.train, cs = "Nigeria", main = "Nigeria",
         var = "Illegal.Carcasses", ylim = c(-2,1))
```

```
## Warning in amcheck(x = x, m = m, idvars = numopts$idvars, priors = priors, : The log transformation i
##  variables with negative values. The values
##  will be shifted up by 1 plus the minimum value
##  of that variable.
```




```r
# PLOT DENSITIES FOR HIGH MISSINGNESS VARIABLES
par(mfrow=c(2,2))
plot(elephants.out.train, which.vars=c(20,42,45,46))
```


Observed and Imputed values of PCPIP
served and Imputed values of Definite.Po
Relative Density
PCPIPCH -- Fraction Missing: 0.006
Definite.Possible -- Fraction Missing: 0.772
served and Imputed values of Illegal.Car
bserved and Imputed values of Percent.Il
Illegal.Carcasses -- Fraction Missing: 0.365
Percent.Illegal -- Fraction Missing: 0.365