Supplemental Material

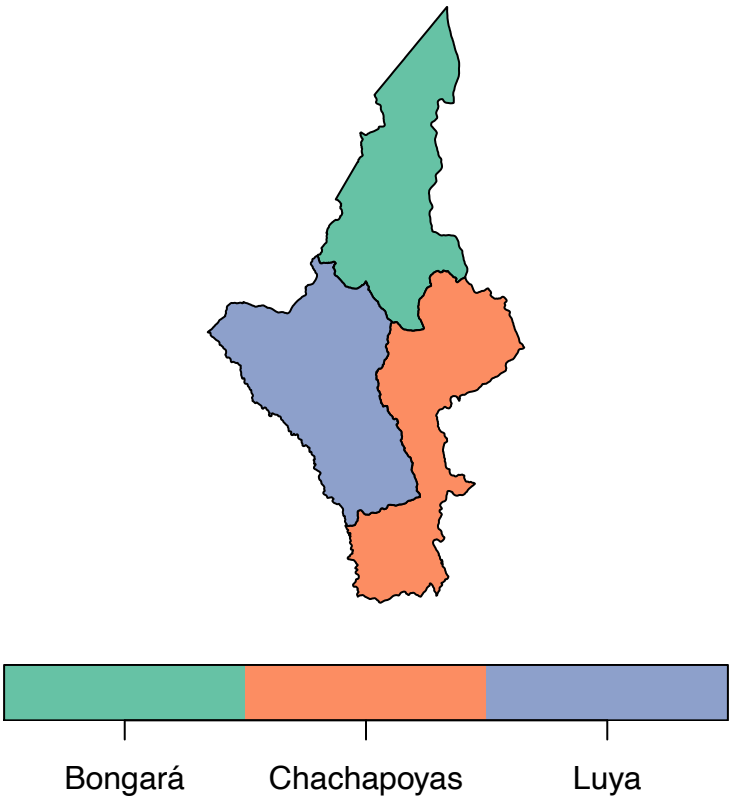
Linking Past and Present Land-use Histories in Southern Amazonas, Peru

07 April, 2021

Define survey area

We make use of GADM vector data to define the extent of the study area and subsequently clip and mask other data files.

```
library(rgdal)
library(tidyverse)
library(sf)
library(raster)

#OSF data files
terrace_pt_OSF <-"https://osf.io/etxhq/download"
nl_poly_OSF <- "https://osf.io/xhu7n/download"
sample_poly_OSF <- "https://osf.io/aujc6/download"

#create mask of three Amazonas provinces used in study
#using GADM vector data
mask <- st_as_sf(getData(name = "GADM", country = "PER", level = 2)) %>%
  filter(NAME_1 == "Amazonas") %>%
  filter(NAME_2 == "Bongará" |
           NAME_2 == "Chachapoyas" |
           NAME_2 == "Luya") %>%
  st_transform(4326)

#calculate their area, in sq km
mask$area <- mask %>%
  st_area() %>%
  units::set_units(value = km^2) %>%
  as.numeric()

plot(mask[7], key.pos = 1, main = "")
```



The total area of the study region is 9187 square kilometers.

Create environmental datasets

Elevation and slope are calculated from 90 meter resolution SRTM data, while distance is calculated using OpenStreetMap data.

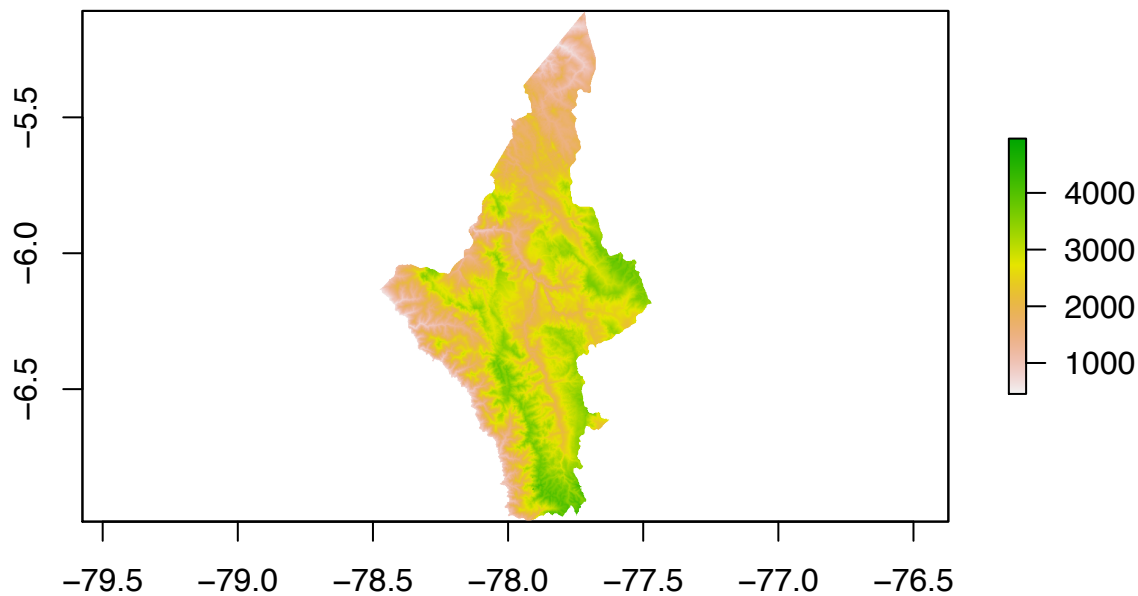
```
#download, crop, and mask DEM data
DEM <- getData("SRTM", lon = -78, lat = -6) %>%
  crop(extent(mask)) %>%
  raster::mask(mask)

#calculate slope from DEM
Slope <- terrain(DEM,
                 opt = "slope",
                 unit = "degrees")

#download and query OpenStreetMap data
library(osmdata)
```

```
## Data (c) OpenStreetMap contributors, ODbL 1.0. https://www.openstreetmap.org/copyright
```

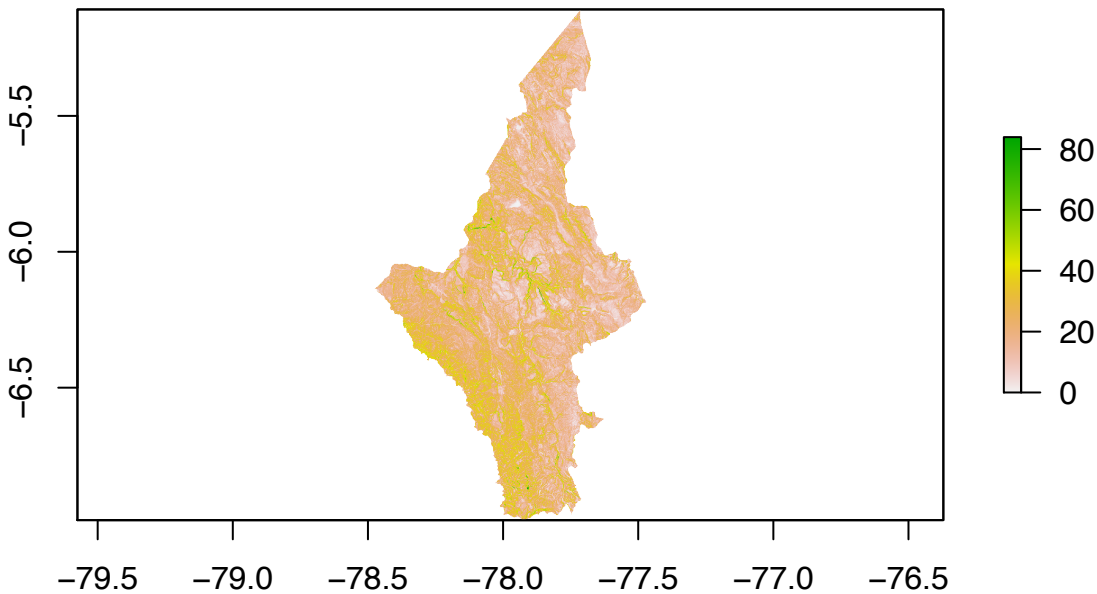
```r
places <- opq(as.numeric(st_bbox(mask))) %>%
  add_osm_feature(key = 'place', value = c("city", "village", "town", "hamlet")) %>%
  osmdata_sf() %>%
  trim_osmdata(summarise(mask))
```

```
## although coordinates are longitude/latitude, st_union assumes that they are planar
```

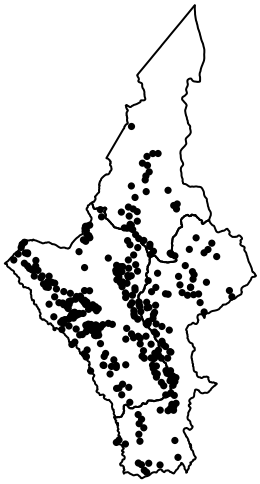
```r
places <- places$osm_points

#plot DEM
plot(DEM)
```




```r
#plot Slope
plot(Slope)
```




```r
#plot settlements
plot(st_geometry(mask))
plot(st_geometry(places), pch = 16, cex = 0.5, add = T)
```



Question 1 - Where did agricultural land-use take place in the past?

Import GeoPACHA data

```
#read in GeoPACHA terrace data
terrace_pt <-terrace_pt_OSF %>%
  download.file("terrace_pt.zip")
unzip("terrace_pt.zip")
terrace_pt <- st_read("terrace_pt.shp")
```

```
## Reading layer `terrace_pt' from data source `/Users/danplekhov/Desktop/terrace_pt.shp' using driver
## Simple feature collection with 2968 features and 6 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: -78.39397 ymin: -6.939521 xmax: -77.51436 ymax: -5.722837
## geographic CRS: WGS 84
```

```
#plot
plot(st_geometry(mask))
plot(st_geometry(terrace_pt), pch = 16, cex = 0.25, add = T)
```

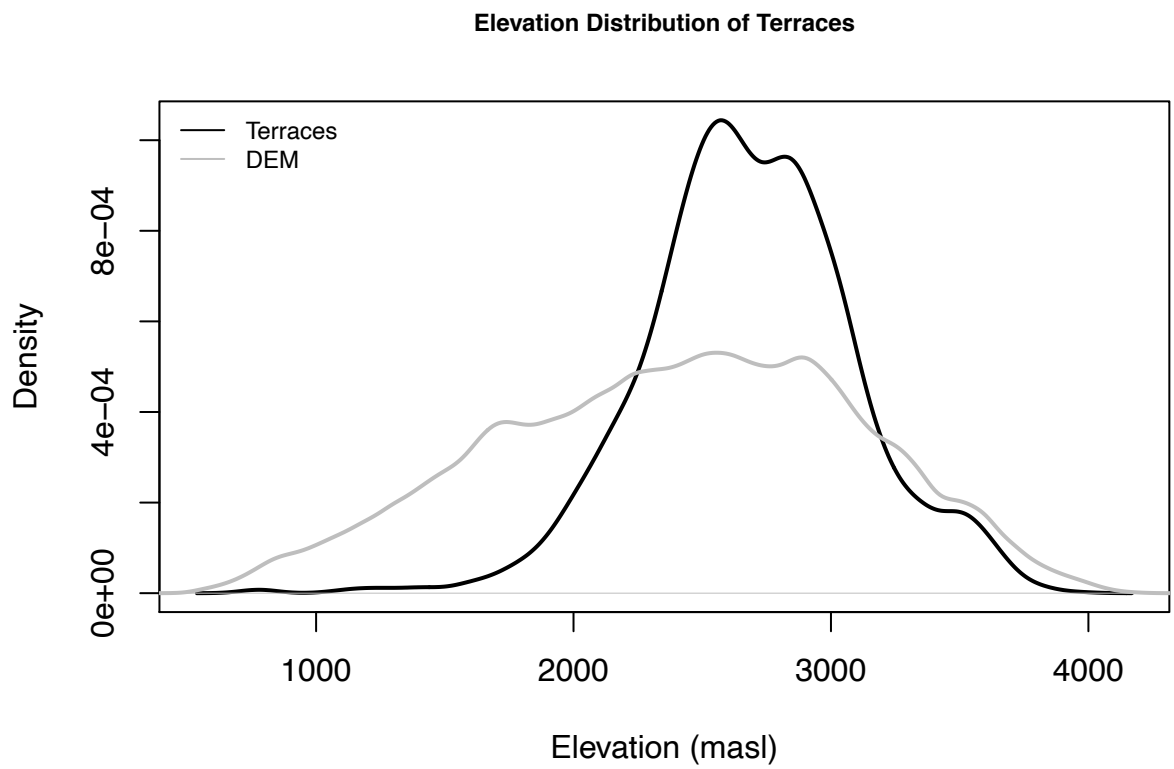
Our GeoPACHA dataset is made up of 2968 points, representing the same number of 500x500 meter grid squares containing agricultural terraces.

Calculate attributes

```
#extract elevation values for each terrace location
terrace_pt$Z <- as.numeric(raster::extract(DEM, terrace_pt))

#extract slope values for each terrace location
terrace_pt$Slope <- as.numeric(raster::extract(Slope, terrace_pt))

#calculate distances from each terrace location to nearest settlement
terrace_pt$Distance <- sapply(1:nrow(terrace_pt), function(i){
  min(st_distance(terrace_pt[i,], places))
  }
  )
```

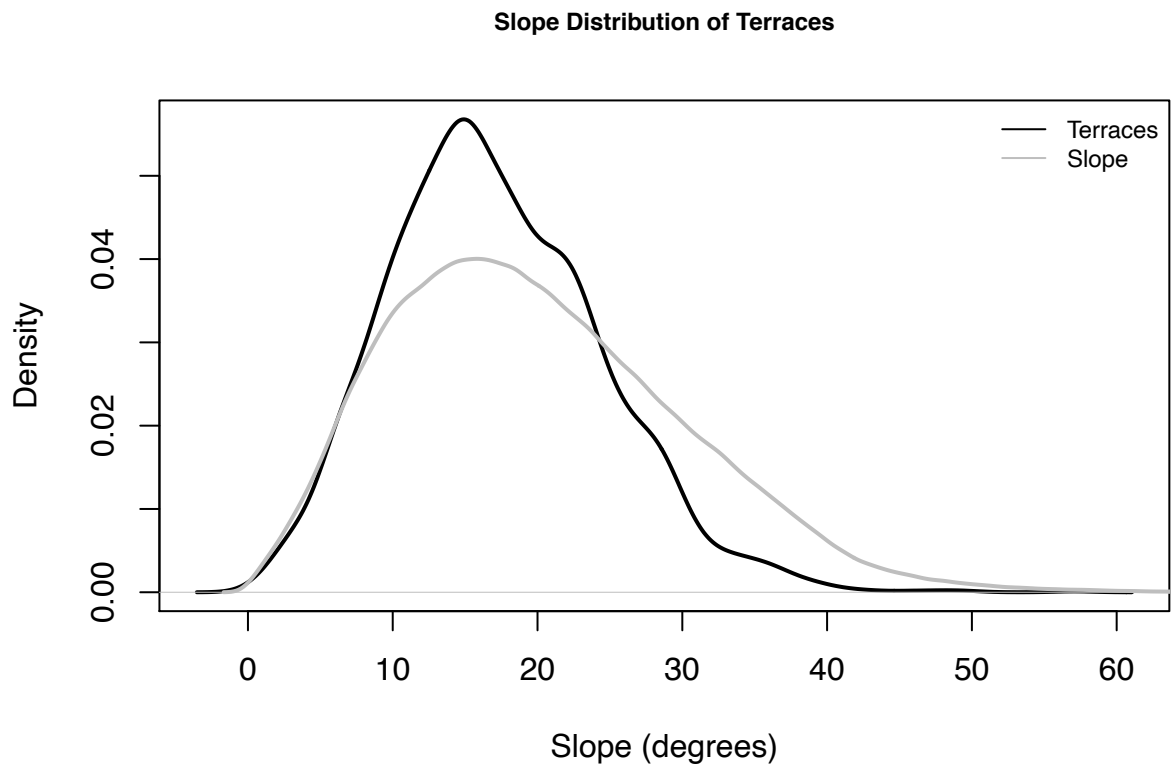
With these data, we can quantitatively and visually assess the environmental and geographic characteristics of our sample of terraces. Beginning with elevation, the average elevation at which terraces in the Chachapoyas region are located is 2684 masl, with a standard deviation of 408 meters. Visual inspection of the density distribution of these elevation values shows an apparent non-normal distribution, which an additional Shapiro-Wilk's test confirms ($W = 0.99$, $p < 0.01$).

```r
plot(density(terrace_pt$Z, na.rm = T),
     main = "Elevation Distribution of Terraces",
     xlab = "Elevation (masl)",
     lwd = 2,
     cex.main=.75)
lines(density(DEM@data@values, na.rm = T),
      col = "grey",
      lwd = 2)
legend("topleft", legend=c("Terraces", "DEM"),
       col=c("black", "grey"), lty=1, cex=0.75, bty = "n")
```



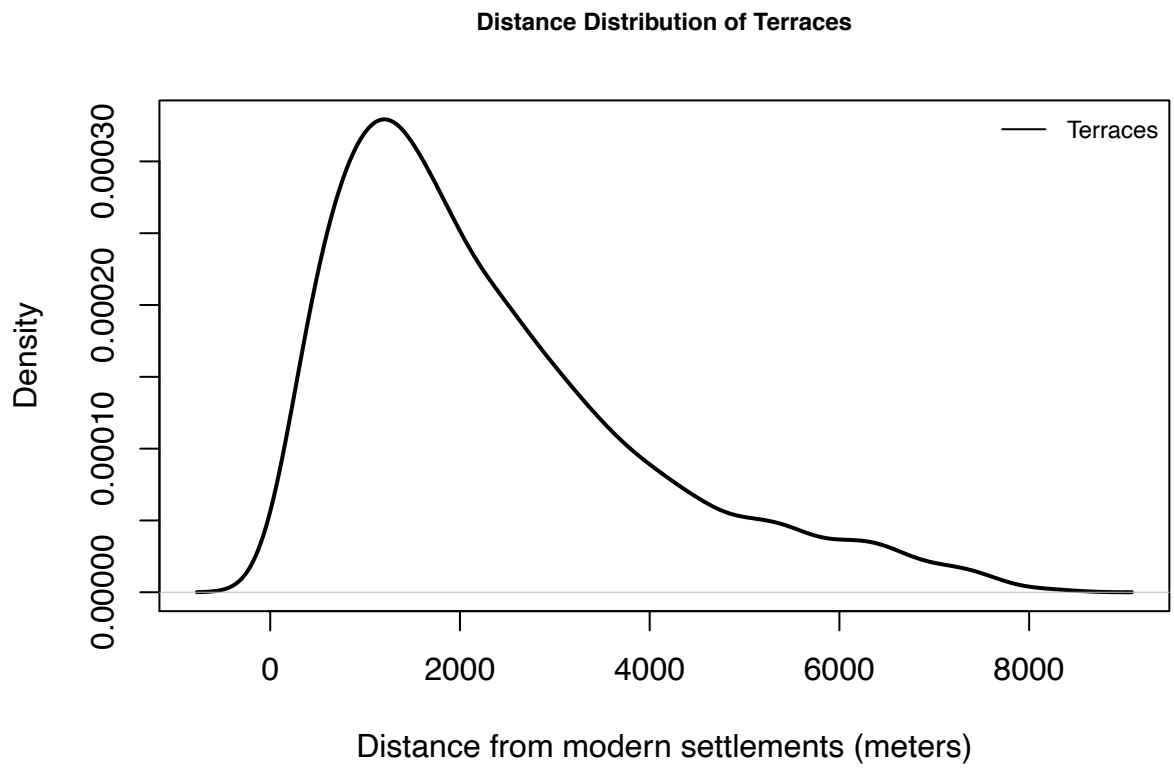
Doing the same for slope, we find that the average slope at which terraces are found in the Chachapoyas region is 17 degrees, with a standard deviation of 7 degrees. The density distribution plot of these slope values shows that this distribution is again not normal ($W = 0.98$, $p < 0.01$).

```r
plot(density(terrace_pt$Slope, na.rm = T),
     main = "Slope Distribution of Terraces",
     xlab = "Slope (degrees)",
     lwd = 2,
     cex.main=.75)
lines(density(Slope@data@values, na.rm = T),
      col = "grey",
      lwd = 2)
legend("topright", legend=c("Terraces", "Slope"),
       col=c("black", "grey"), lty=1, cex=0.75, bty = "n")
```



Finally, looking at distance to settlements, we find that the average Euclidean distance terraces are found from settlements in the Chachapoyas region is 2332 meters, with a standard deviation of 1637 meters. The density distribution plot of these distance values shows that this distribution is again not normal (W = 0.91, $p < 0.01$).

```
plot(density(terrace_pt$Distance, na.rm = T),
     main = "Distance Distribution of Terraces",
     xlab = "Distance from modern settlements (meters)",
     lwd = 2,
     cex.main=.75)
legend("topright", legend=c("Terraces"),
       col=c("black"), lty=1, cex=0.75, bty = "n")
```




```
## pdf
##   2
```

That the distribution of all three variables is non-normal indicates that they are being constrained or structured in a way that is not random.

Question 2 - Where is agricultural land-use taking place today?

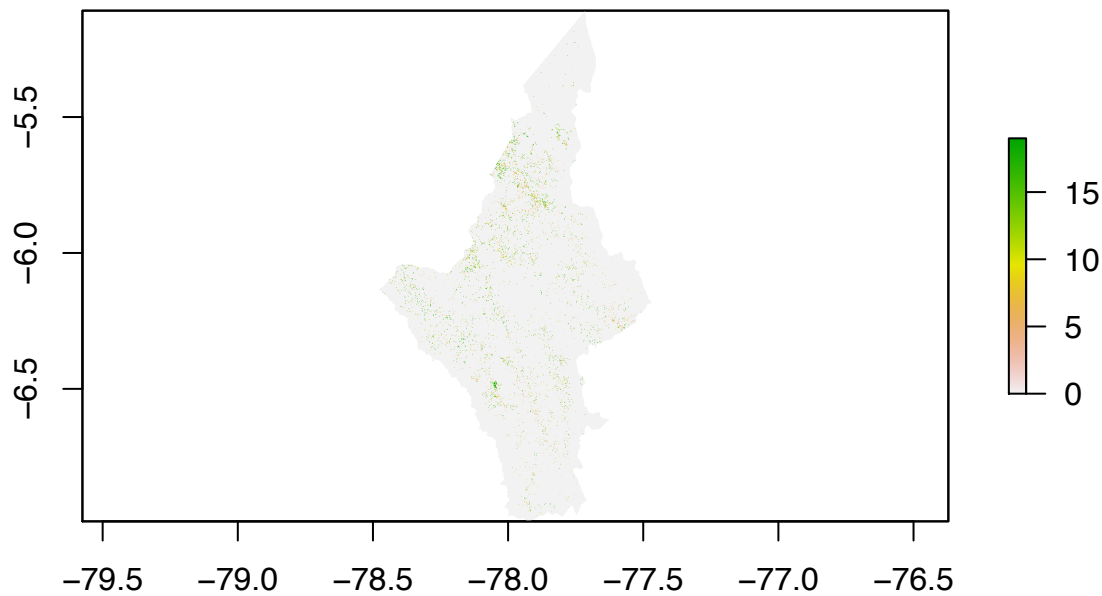
Import forest loss data

The source for our forest loss data comes from the Global Land Analysis & Discovery laboratory at the University of Maryland, specifically their publicly and freely available Global Forest Change 2000–2019 data. The tile used in this study is the lossyear granule with top-left corner at 0N, 80W. Lossyear data covers the period 2000–2019 and is "defined as a stand-replacement disturbance, or a change from a forest to non-forest state. Encoded as either 0 (no loss) or else a value in the range 1–19, representing loss detected primarily in the year 2001–2019, respectively."

```
#download forest loss raster and mask to study area
fl_filepath <- "https://storage.googleapis.com/earthenginepartners-hansen/GFC-2019-v1.7/Hansen_GFC-2019-

fl <- fl_filepath %>%
  download.file("lossyear_00N_080W.tif")
fl <- raster("lossyear_00N_080W.tif") %>%
```

```
  crop(extent(mask)) %>%
  mask(mask)
```

```
plot(fl)
```



Summary of Data

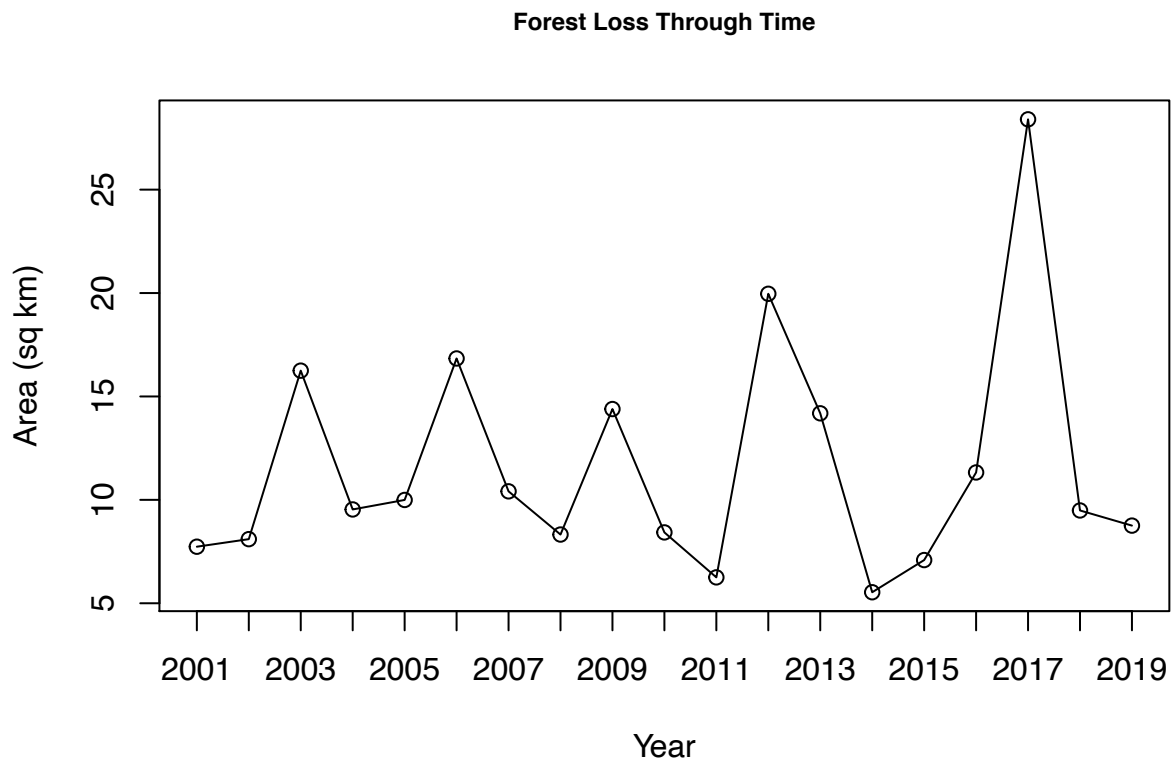
Having imported and masked the forest loss data (hereafter in kobject names: fl), we calculate some summary statistics and generate plots to explore some initial trends in forest loss within our study area.

```
#classify values according to year of forest loss, or no loss
years <- c(0, seq(2001, 2019, 1))

#count number of cells with each value
fl_freq <- as_tibble(freq(fl, useNA = "no"))
fl_freq$years <- years

#determine resolution of projected forest loss data in sq meters
fl_res <- res(projectRaster(fl, crs = CRS("+init=EPSG:32718")))[1]^2

#calculate total area of forest loss for each year, in sq km
fl_freq$area <- (fl_freq$count * fl_res)/1000000

#calculate percentage of total area lost per year
```

```r
fl_freq$percent <- (fl_freq$area/sum(mask$area))*100

knitr::kable(fl_freq, format="markdown", digits = 2)
```

value	count	years	area	percent
0	11720974	0	8993.39	97.90
1	10080	2001	7.73	0.08
2	10557	2002	8.10	0.09
3	21176	2003	16.25	0.18
4	12429	2004	9.54	0.10
5	13027	2005	10.00	0.11
6	21944	2006	16.84	0.18
7	13570	2007	10.41	0.11
8	10847	2008	8.32	0.09
9	18755	2009	14.39	0.16
10	10981	2010	8.43	0.09
11	8148	2011	6.25	0.07
12	26019	2012	19.96	0.22
13	18487	2013	14.18	0.15
14	7211	2014	5.53	0.06
15	9238	2015	7.09	0.08
16	14760	2016	11.33	0.12
17	37012	2017	28.40	0.31
18	12364	2018	9.49	0.10
19	11410	2019	8.75	0.10

```r
#show change over time of forest loss
plot(fl_freq$area[2:20],
     type = "o",
     main = "Forest Loss Through Time",
     xlab = "Year",
     ylab = "Area (sq km)",
     xaxt = "n",
     cex.main = 0.75)
axis(1,
     at = fl_freq$value[2:20],
     labels = fl_freq$years[2:20])
```



We find that the total area of forest loss from 2001 to 2019 is 221 square kilometers. As the plot above shows, losses are not consistent from year to year but rather fluctuate around an average annual loss of 11.63 square kilometers.

Convert to Polygons

For the next step of the analysis, we convert the `fl` raster to two polygon vectors, one representing areas with no forest loss (`nl_poly`) and the other representing only those areas with forest loss (`fl_poly`). For both, we aggregate contiguous cells of equal values.

```
library(stars)

#recode areas in forest loss raster with no forest loss as NA
fl[fl == 0] <- NA

#convert raster to polygons, merging adjacent
#cells with identical values
fl_poly <- fl %>%
  st_as_stars() %>%
  st_as_sf(na.rm = T, as_points = F, merge = T)

#name columns
names(fl_poly) <- c("value", "geometry")
```

```r
#create new raster (nl) with only 0 values (no loss)
nl <- fl
nl[!nl == 0] <- NA

#convert raster to polygons, merging adjacent
#cells with identical values
nl_poly <- nl %>%
  st_as_stars() %>%
  st_as_sf(na.rm = T, as_points = F, merge = T) %>%
  st_buffer(0) %>%
  summarise(value = 0) %>%
  st_cast("MULTIPOLYGON")

#name columns
names(nl_poly) <- c("value", "geometry")
```

Now vectorized, we can again calculate some summary statistics, though now based on the CFLAs and NFLAs rather than individual pixels.

```r
#calculate area for contiguous forest loss areas (CFLAs)
fl_poly$area <- fl_poly %>%
  st_area() %>%
  units::set_units(value = m^2) %>%
  as.numeric()

#summary statistics for forest loss areas by year
summary(fl_poly$area)
```

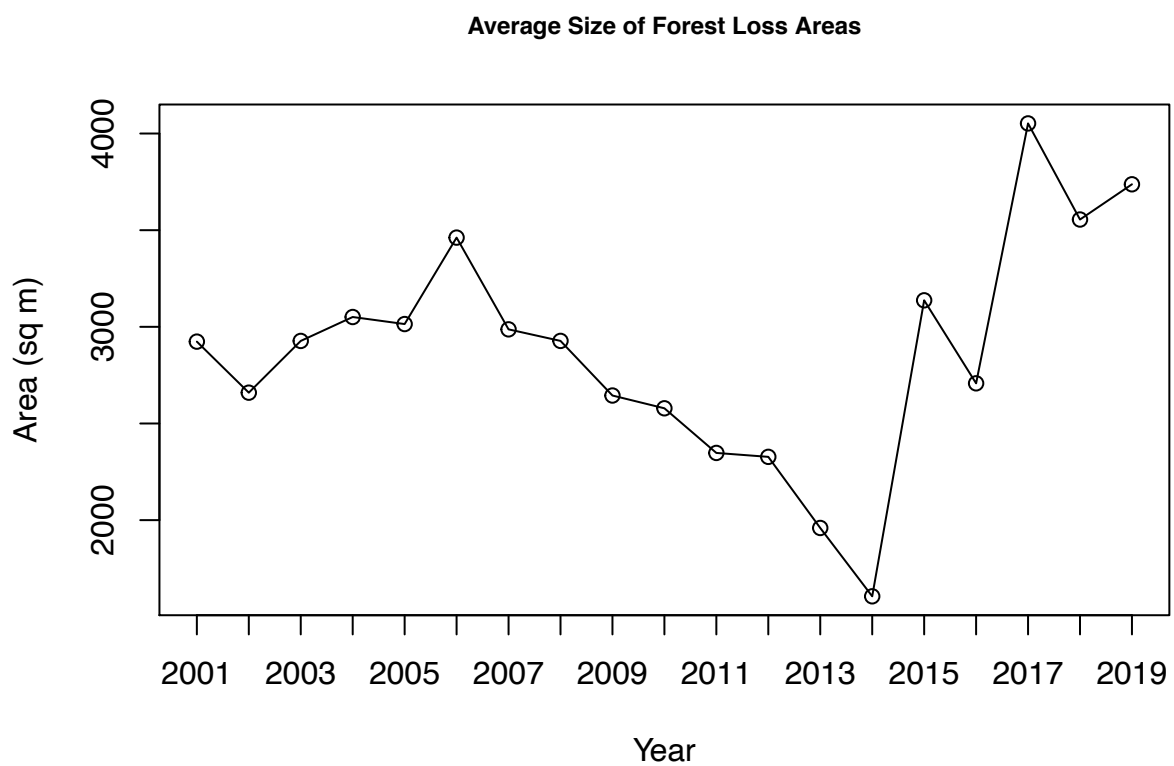
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##     763.8     765.0    1528.4    2834.2    2297.2 2124621.6
```

```r
#calculate summary statistics per year
fl_summary <- st_drop_geometry(fl_poly) %>%
  group_by(value) %>%
  summarize("Number of CFLAs" = n(), "Average area (sq m)" = mean(area))

#print table
knitr::kable(fl_summary, format="markdown", digits = 2)
```

value	Number of CFLAs	Average area (sq m)
1	2638	2923.40
2	3037	2659.36
3	5534	2927.10
4	3117	3050.73
5	3306	3014.41
6	4850	3461.64
7	3476	2986.89
8	2835	2927.20
9	5425	2644.69
10	3258	2578.57
11	2655	2347.68

value	Number of CFLAs	Average area (sq m)
12	8553	2327.43
13	7218	1959.54
14	3435	1605.97
15	2253	3137.10
16	4170	2707.71
17	6988	4052.23
18	2661	3555.06
19	2336	3737.33

```r
#plot average size of CFLA through time
plot(fl_summary$`Average area (sq m)`,
     type = "o",
     main = "Average Size of Forest Loss Areas",
     xlab = "Year",
     ylab = "Area (sq m)",
     xaxt = "n",
     cex.main = .75)
axis(1,
     at = fl_summary$value,
     labels = years[2:20])
```



```r
#summary statistics for no loss areas
nl_poly$area <- nl_poly %>%
  st_area() %>%
  units::set_units(value = km^2) %>%
  as.numeric()
```

```
## pdf
##   2
```

We now begin the process of subsetting our data, which we do first by selecting only the fourth quantile of CFLAs according to area, thereby excluding CFLAs made up of only one or two cells in the original forest loss raster.

```r
#quantile of CFLAs by area
fl_quantile <- quantile(fl_poly$area)

#quantile breakdown of area
fl_quantile
```

```
##           0%          25%          50%          75%         100%
##     763.7574     764.9637    1528.4329    2297.1532 2124621.6279
```

```r
#subset fourth quantile forest loss areas
fl_poly_subset <- fl_poly[fl_poly$area>fl_quantile[4],]

#summary statistics for area
summary(fl_poly_subset$area)
```

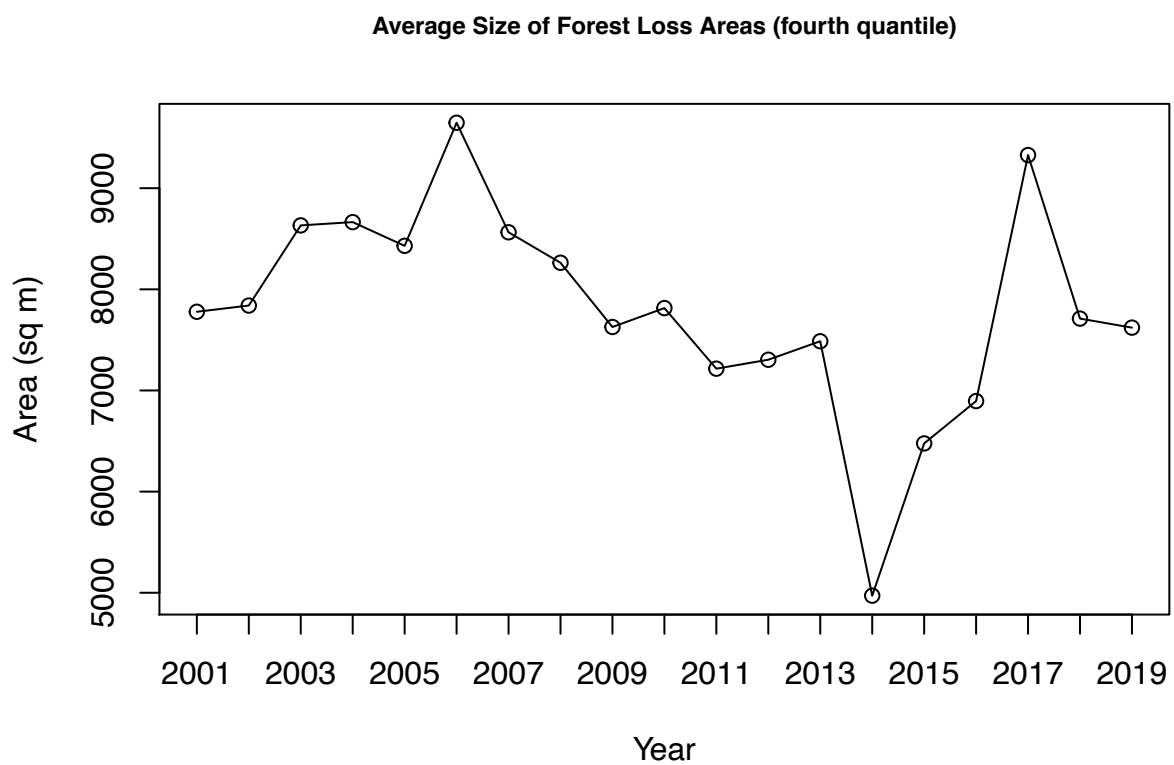
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2297    3823    5354    8022    9173 2124622
```

```r
fl_summary_subset <- st_drop_geometry(fl_poly_subset) %>%
  group_by(value) %>%
  summarize("Number of CFLAs" = n(), "Average area (sq m)" = mean(area))

knitr::kable(fl_summary_subset, format="markdown", digits = 2)
```

value	Number of CFLAs	Average area (sq m)
1	717	7778.11
2	709	7839.63
3	1344	8632.16
4	804	8664.73
5	866	8430.25
6	1335	9646.97
7	874	8563.84
8	721	8262.75
9	1275	7627.69
10	716	7814.65
11	545	7215.44
12	1704	7303.80

value	Number of CFLAs	Average area (sq m)
13	1094	7486.42
14	462	4971.23
15	758	6476.40
16	1135	6895.26
17	2473	9327.46
18	974	7710.54
19	930	7621.41

```r
#plot of CFLA areas (only fourth quantile)
plot(fl_summary_subset$"Average area (sq m)",
     type = "o",
     main = "Average Size of Forest Loss Areas (fourth quantile)",
     xlab = "Year",
     ylab = "Area (sq m)",
     xaxt = "n",
     cex.main = .75)
axis(1,
     at = fl_summary_subset$value,
     labels = years[2:20])
```



This leads to a subsequent reduction in total CFLA numbers from 77745 to 19436.

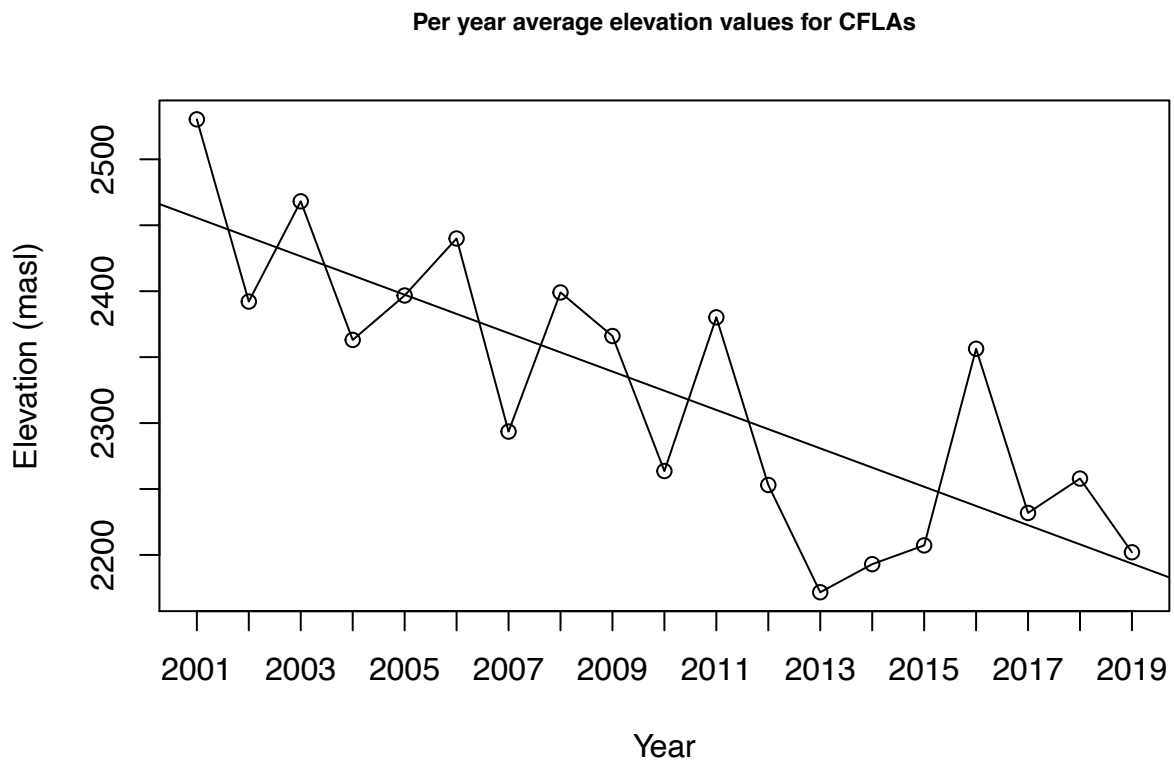
Calculate attributes

It is at this stage that we can calculate the distribution of elevation, slope, and distance values for these forest loss areas and look at inter-annual trends.We convert the forest loss polygons to points, using the centroid of each polygon. We do so to accelerate processing and to make the forest loss data more comparable to our terrace data.

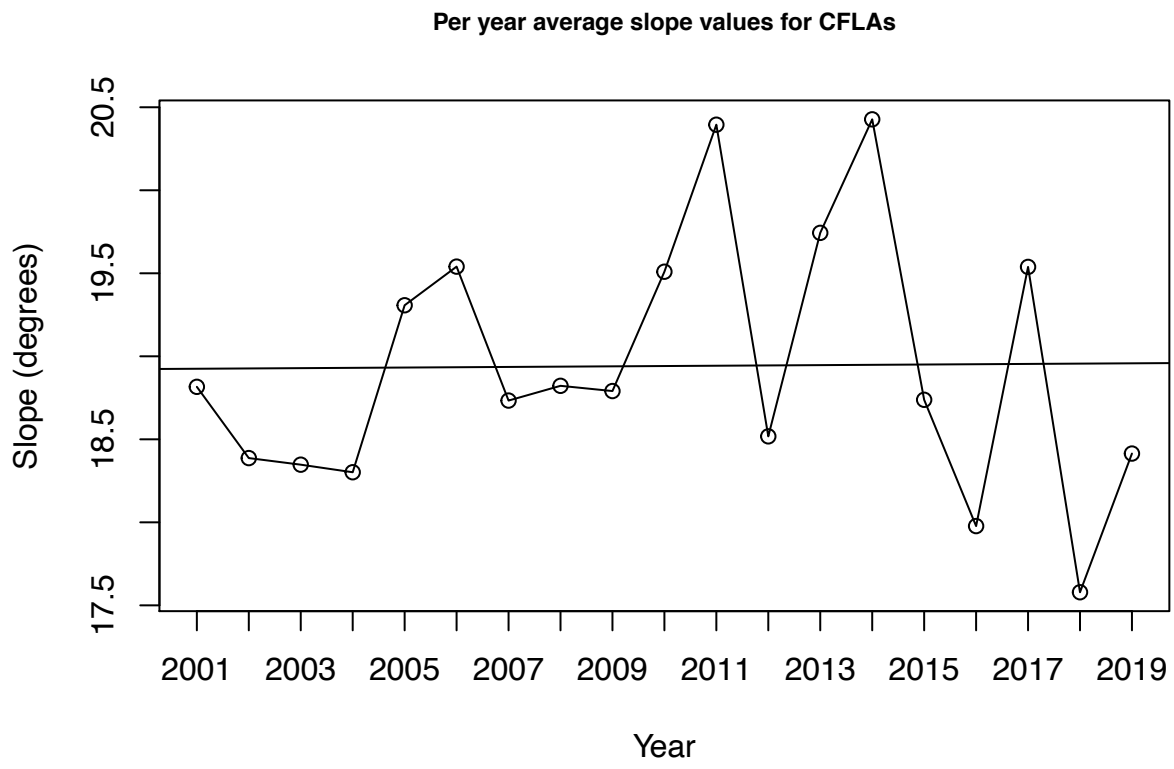
```
#create new spatial object using centroids of CFLAs
fl_point <- st_centroid(fl_poly_subset)

#extract elevation values for each CFLA
fl_point$Z <- as.numeric(raster::extract(DEM, fl_point))
#extract slope values for each CFLA
fl_point$Slope <- as.numeric(raster::extract(Slope, fl_point))
#calculate distances from each CFLA to nearest settlement
fl_point$Distance <- sapply(1:nrow(fl_point), function(i){
  min(st_distance(fl_point[i,], places))
  }
  )
```

We find that the average elevation at which forest loss is taking place is 2319 masl, with a standard deviation of 526 masl. The average slope is 19 degrees, with a standard deviation of 9. The average distance to settlements is 3613 meters, with a standard deviation of 3835.

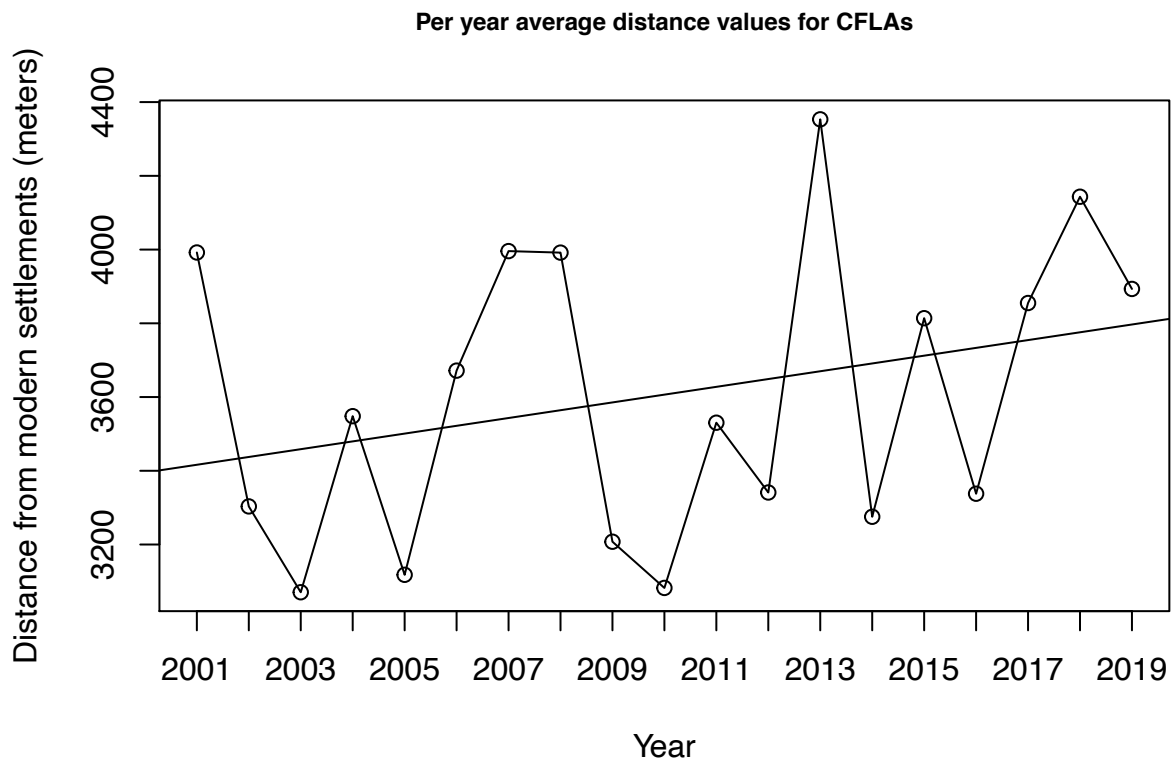
```
#average values per year
fl_summary_point <- st_drop_geometry(fl_point) %>%
  group_by(value) %>%
  summarize("Elevation" = mean(Z, na.rm = T),
            "Slope" = mean(Slope, na.rm = T),
            "Distance" = mean(Distance))

#elevation
plot(fl_summary_point$Elevation,
     type = "o",
     main = "Per year average elevation values for CFLAs",
     xlab = "Year",
     ylab = "Elevation (masl)",
     xaxt = "n",
     cex.main = .75)
axis(1,at = fl_summary_point$value,
     labels = years[2:20])
abline(lm(Elevation~value, data = fl_summary_point))
```




```r
#slope
plot(fl_summary_point$Slope,
     type = "o",
     main = "Per year average slope values for CFLAs",
     xlab = "Year",
     ylab = "Slope (degrees)",
     xaxt = "n",
     cex.main = .75)
axis(1,at = fl_summary_point$value,
     labels = years[2:20])
abline(lm(Slope~value, data = fl_summary_point))
```




```r
#distance
plot(fl_summary_point$Distance,
     type = "o",
     main = "Per year average distance values for CFLAs",
     xlab = "Year",
     ylab = "Distance from modern settlements (meters)",
     xaxt = "n",
     cex.main = .75)
axis(1,at = fl_summary_point$value,
     labels = years[2:20])
abline(lm(Distance~value, data = fl_summary_point))
```




```
## pdf
##   2
```

Displayed visually, we note a clear negative relationship between elevation and year of deforestation and a more modest positive relationship between distance and year of deforestation. There is no apparent relationship between slope and year of deforestation.

Sampling

At this stage, we move to the generation of random samples within forest loss and no forest loss areas to investigate the presence/absence of terracing therein. Due to limitations in publicly available high-resolution satellite imagery for studying ancient terrace locations, we limit further analysis of forest loss areas to the years 2001-2016.

```
#exclude areas with forest loss after 2016
fl_poly_subset <- fl_poly_subset[fl_poly_subset$value <= 16,]
```

This results in a reduction in total CFLA numbers from 77745 to 15059. Of these CFLAs, we take a 10% sample (n = 1505.9) and divide it evenly among the investigators. Each investigator then analyzed their respective sample of CFLAs, noting whether an area had terraces present (YES) or not (NO). A third category (MAYBE) is assigned to those CFLAs for which it is impossible to determine the presence of terraces, due to either insufficient imagery, cloud cover, or subsequent reforestation.

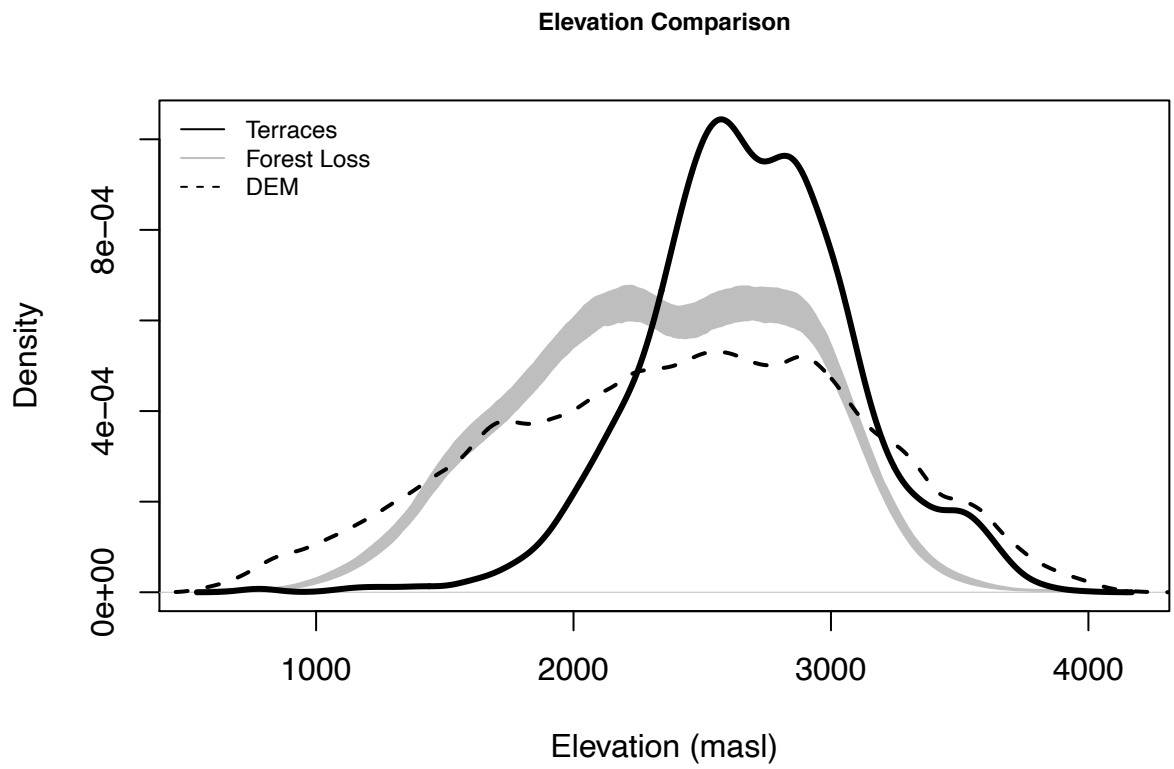
Additionally, we take random samples of the forest loss data equal in number to our terrace data (n = 2968). Because of the large size of the forest loss dataset, we can take numerous random samples of these data to

calculate confidence intervals for comparing the distribution of these forest loss data to that of the ancient terraces.

Question 3 - Does ancient land-use have an impact on modern land-use?

To evaluate this question, we can begin by simply comparing distributions and averages.

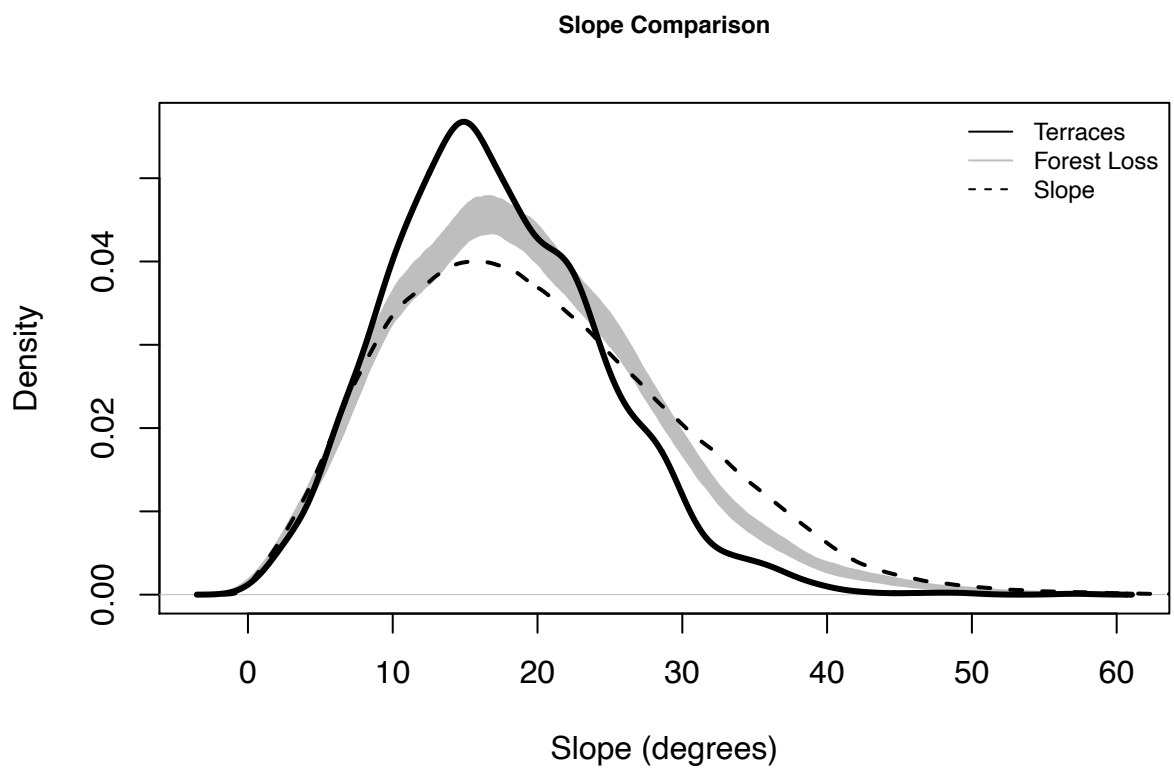
```
#exclude years after 2016 from fl_point
fl_point_subset <- fl_point[fl_point$value <= 16,]

#compare density distribution for elevation
fit1 <- density(terrace_pt$Z, na.rm = T)
fit2 <- replicate(500,{
  fl_point_sample <- sample_n(fl_point_subset,
                              nrow(terrace_pt),
                              replace = F)
  density(fl_point_sample$Z, from=min(fit1$x), to=max(fit1$x), na.rm = T)$y
})
fit3 <- apply(fit2, 1, quantile, c(0.025,0.975) )

#create plot
plot(fit1,
     main = "Elevation Comparison",
     xlab = "Elevation (masl)",
     cex.main = 0.75)
polygon( c(fit1$x, rev(fit1$x)), c(fit3[1,], rev(fit3[2,])),
        col='grey', density = -0.5, border=F)
lines(fit1,
      col = "black",
      lwd = 3)
lines(density(DEM@data@values, na.rm = T),
      col = "black",
      lwd = 2,
      lty = 2)
legend("topleft",
       legend=c("Terraces", "Forest Loss", "DEM"),
       col=c("black", "grey", "black"),
       lty=c(1,1,2),
       cex=0.75,
       bty = "n")
```



Thus, we find that ancient terraces are found, on average, 340 meters higher in elevation than modern-day forest loss. Visual inspection of these density distributions shows that while they are overlapping, ancient terrace elevation values have a more restricted distribution than that of forest loss. We determine that these distributions are statistically different through a Kolmogorov-Smirnov test (D = 0.34, p = 0)

```
#compare density distribution for slope
fit1 <- density(terrace_pt$Slope, na.rm = T)
fit2 <- replicate(500,{
  fl_point_sample <- sample_n(fl_point_subset,
                              nrow(terrace_pt),
                              replace = F)
  density(fl_point_sample$Slope, from=min(fit1$x), to=max(fit1$x), na.rm = T)$y
})
fit3 <- apply(fit2, 1, quantile, c(0.025,0.975) )

#create plot
plot(fit1,
     main = "Slope Comparison",
     xlab = "Slope (degrees)",
     cex.main = 0.75)
polygon( c(fit1$x, rev(fit1$x)), c(fit3[1,], rev(fit3[2,])),
        col='grey', density = -0.5, border=F)
lines(fit1,
      col = "black",
      lwd = 3)
lines(density(Slope@data@values, na.rm = T),
```

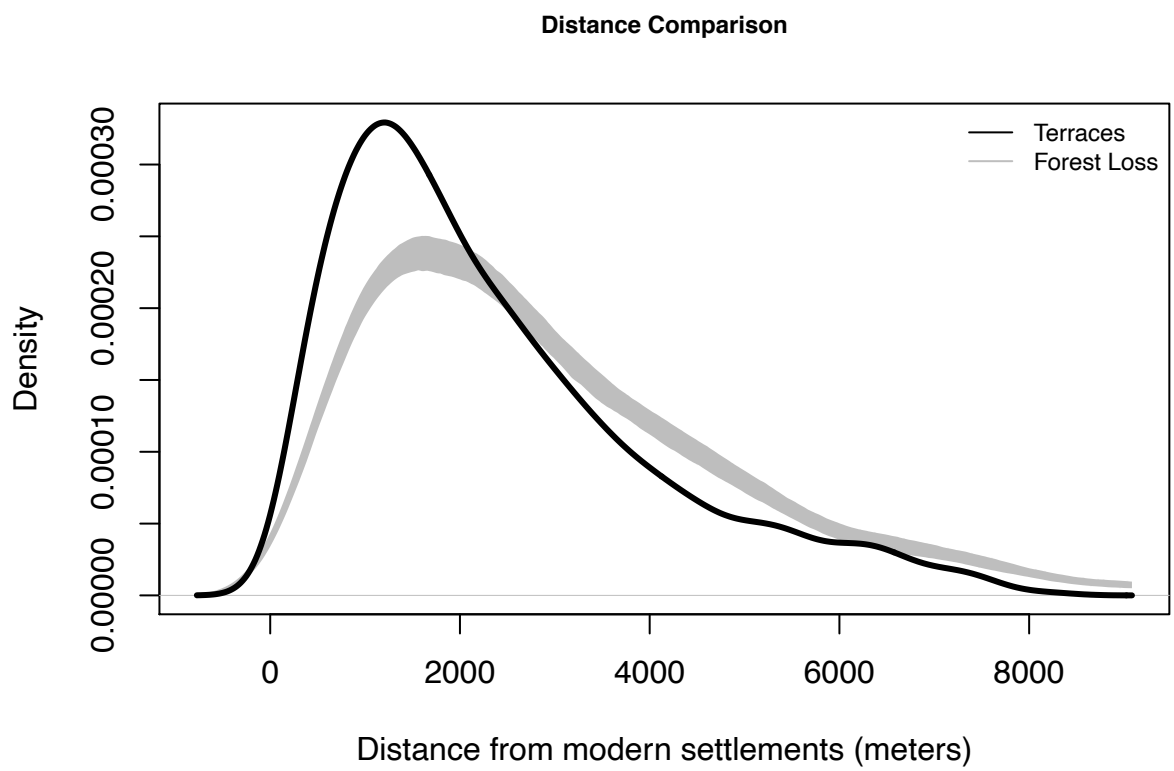
```
      col = "black",
      lwd = 2,
      lty = 2)
legend("topright",
       legend=c("Terraces", "Forest Loss", "Slope"),
       col=c("black", "grey","black"),
       lty=c(1,1,2),
       cex=0.75,
       bty = "n")
```



For slope, we find that, on average, ancient terraces around found on slopes about 2 degrees less steep than where modern-day forest loss is taking place. Though these distributions are far more overlapping than elevation, we find that they are nevertheless statistically distinct (D = 0.1, p = 0)

```
#compare density distribution for distance
fit1 <- density(terrace_pt$Distance, na.rm = T)
fit2 <- replicate(500,{
  fl_point_sample <- sample_n(fl_point_subset,
                              nrow(terrace_pt),
                              replace = F)
  density(fl_point_sample$Distance, from=min(fit1$x), to=max(fit1$x), na.rm = T)$y
})
fit3 <- apply(fit2, 1, quantile, c(0.025,0.975) )

#create plot
```

```r
plot(fit1,
     main = "Distance Comparison",
     xlab = "Distance from modern settlements (meters)",
     cex.main = 0.75)
polygon( c(fit1$x, rev(fit1$x)), c(fit3[1,], rev(fit3[2,])),
        col='grey', density = -0.5, border=F)
lines(fit1,
      col = "black",
      lwd = 3)
legend("topright",
       legend=c("Terraces", "Forest Loss"),
       col=c("black", "grey"),
       lty=1,
       cex=0.75,
       bty = "n")
```



Finally, for distance, we find that, on average, ancient terraces are found on slopes about 1189 degrees less steep than where modern-day forest loss is taking place. These distributions are statistically different (D = 0.17, p = 0)

```
## pdf
##   2
```

Next, we take a smaller 10% sample of the CFLAs and no loss areas and investigate each to determine whether they contain terraces.

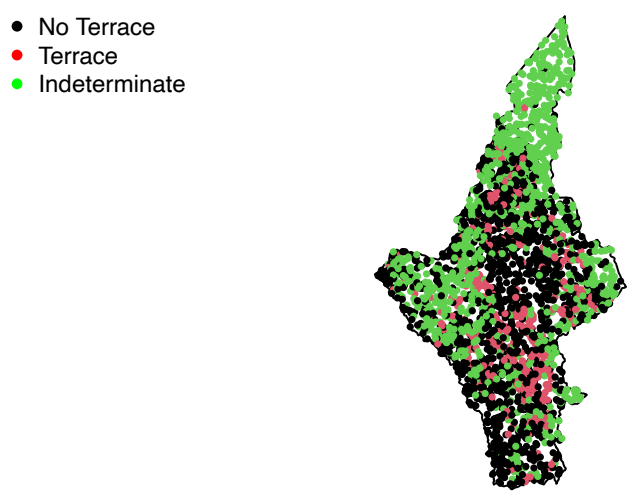
Calculate attributes for sampled CFLAs

For each CFLA, we next calculate its average elevation and slope as well as its distance to the nearest populated place.

```
#extract elevation values
sample$Z <- as.numeric(raster::extract(DEM, sample))
#extract slope values
sample$Slope <- as.numeric(raster::extract(Slope, sample))
#calculate distance values
sample$Distance <- sapply(1:nrow(sample), function(i){
  min(st_distance(sample[i,], places))
  }
  )
```

We check at this stage whether there is any evident spatial patterning in the points coded as "MAYBE" for the presence of terraces, which might indicate that their exclusion leads us to miss some underlying process. Visually, there appears to be no pattern, with the exception of a high density of "MAYBE" points in the northern half of Bongará. This area corresponds to the Zona Reservada Río Nieva, where there is limited development and correspondingly poor satellite imagery coverage.

```
plot(st_geometry(mask))
plot(st_geometry(sample), add = T, pch = 16, cex = 0.5, col = factor(sample$terrace, levels = c("NO", "
legend("topleft", legend=c("No Terrace", "Terrace", "Indeterminate"),
       col=c("black", "red", "green"), pch=16, cex=0.75, bty = "n")
```



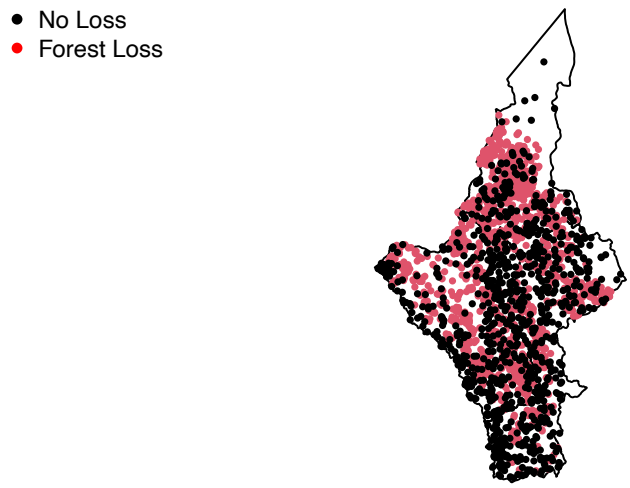
```r
#this is where I tried to find a way to test for autocorrelation and CSR.

library(spatstat)
library(maptools)

#project data for this step of analysis
mask_sp <- mask %>%
  st_transform(32718) %>%
  as("Spatial")

#define window
window <- as.owin.SpatialPolygons(mask_sp)

#project, subset, and convert sample data to ppp
sample_ppp <- sample %>%
  st_transform(32718) %>%
  ppp(x = st_coordinates(.)[,1],
      y = st_coordinates(.)[,2],
      marks = .,
      window = window,
      check = T)

plot(quadrat.test(sample_ppp))
```

We thus exclude those points that are coded as "MAYBE" and recode the data for further analyses.

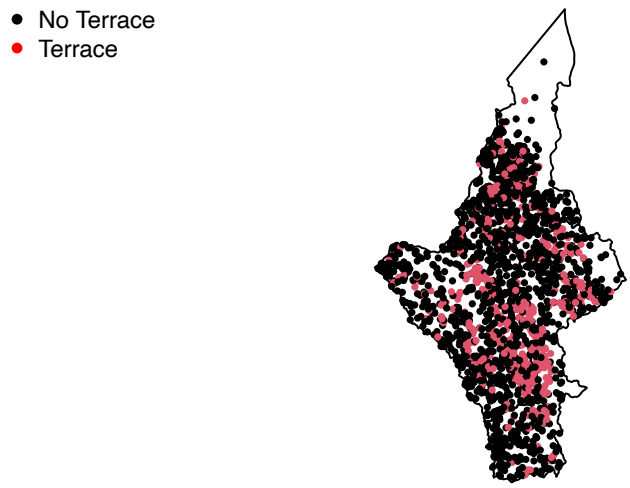
```r
#recode points from MAYBE to NA
sample$terrace[sample$terrace=="MAYBE"] <- NA

#exclude NA points
sample <- sample[-which(is.na(sample$terrace)),]

#recode NO/YES to 0/1
sample$terrace <- factor(sample$terrace,
                     levels = c("NO", "YES"),
                     labels = c(0,1))
sample$terrace <- as.numeric(as.character((sample$terrace)))

#smooth out bad data
sample$Slope[which(is.nan(sample$Slope))] <- mean(sample$Slope, na.rm = T)
sample$Slope[which(is.na(sample$Slope))] <- mean(sample$Slope, na.rm = T)


#create new column corresponding to whether point is in fl or nl area
sample$fl <- ifelse(sample$value == 0, 0, 1)
```

We can visualize the distribution of our coded attributes before proceeding to further analysis.

```r
plot(st_geometry(mask))
plot(st_geometry(sample), add = T, pch = 16, cex = 0.5, col = as.factor(sample$fl))
legend("topleft", legend=c("No Loss", "Forest Loss"),
       col=c("black", "red"), pch=16, cex=0.75, bty = "n")
```




```
plot(st_geometry(mask))
plot(st_geometry(sample), add = T, pch = 16, cex = 0.5, col = as.factor(sample$terrace))
legend("topleft", legend=c("No Terrace", "Terrace"),
       col=c("black", "red"), pch=16, cex=0.75, bty = "n")
```



Exploring causal relationships

To test what effect the presence of terraces has on where forest loss is occuring today, we build a general linear model, with terrace presence, elevation, slope, and distance as explanatory variables and forest loss as our outcome variable. We also conduct ANOVA likelihood ratio tests to determine which variables contribute meaningfully to the model, as determined by sequentially adding variables and testing whether each addition significantly predicts more of the model deviance.

```
#build the model
m1 <- glm(fl ~ terrace + Z + Slope + Distance, data = sample,
          family = binomial,
          na.action = na.exclude)


#Type II so it is order independent
library(car)
library(lmtest)
library(modEvA)

#which variables contribute to the model
Anova(m1, type = "2", test = "LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: fl
```

```
##          LR Chisq Df Pr(>Chisq)
## terrace   141.535  1  < 2.2e-16 ***
## Z          26.700  1  2.376e-07 ***
## Slope      12.041  1  0.0005203 ***
## Distance    0.309  1  0.5782736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#is the model with the variables a significant improvement
#over one in which only mean values are used?
anova(update(m1, ~1), m1, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: fl ~ 1
## Model 2: fl ~ terrace + Z + Slope + Distance
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      2169     2918.4
## 2      2165     2745.5  4   172.96 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of these two anova tests indicate that all our variables, except Distance, significantly improve the predicative power of our model for modeling forest loss.

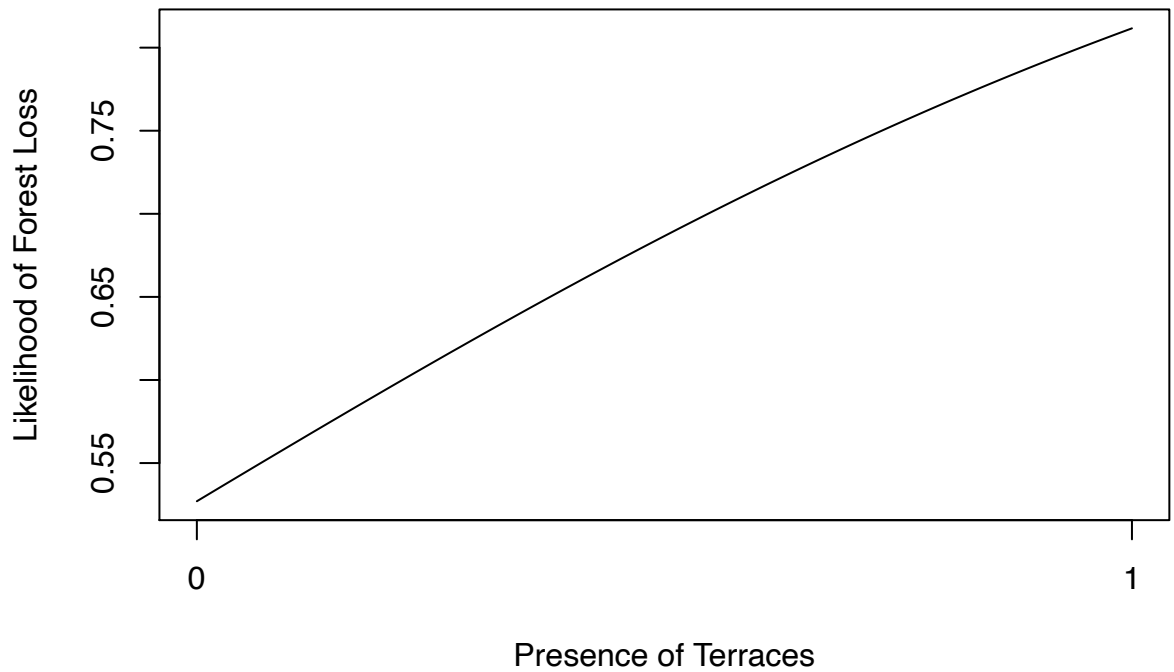
```r
#calculate D-squared value for model
Dsquared(m1)
```

```
## [1] 0.05926606
```

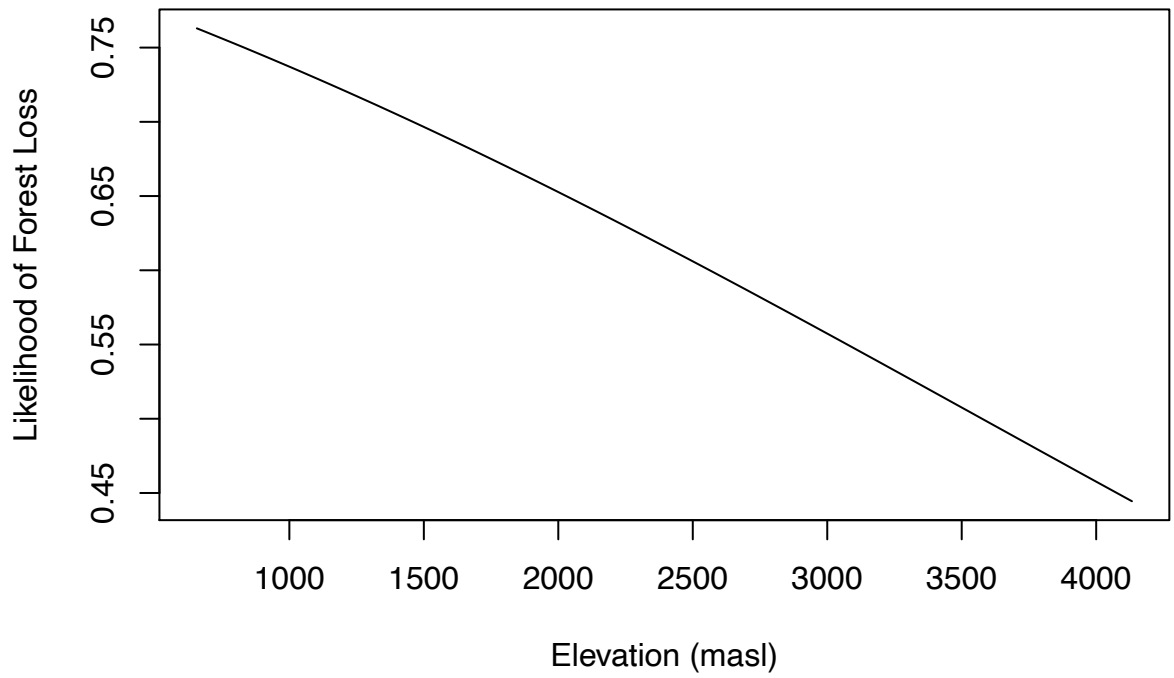
Even so, the overall amount of variance that our model is able to predict is about 6 percent. We can therefore say that, while the presence of terraces does have a significant impact on whether forest loss takes place or not, forest loss does not seem to be influenced in any meaningful way by the distribution of terraces.

We can visualize this effect with a plot showing how the likelihood that an area experiences forest loss changes with and without terraces, while holding our other variables constant.

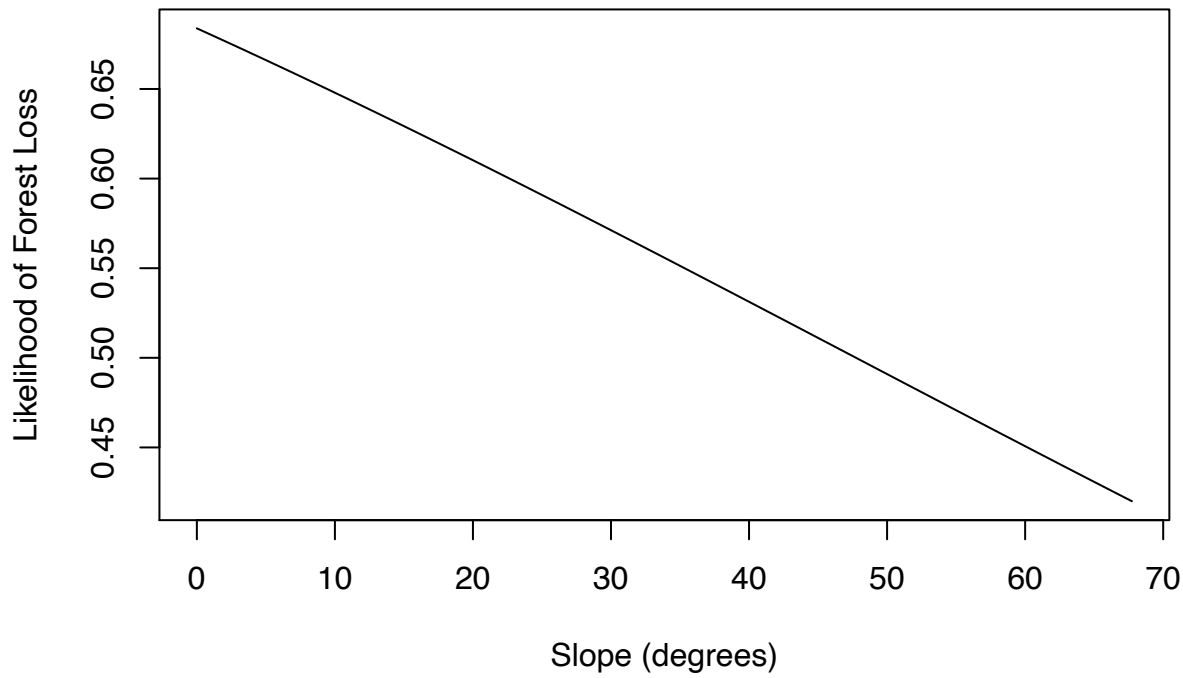
```r
#create new data frame of possible terrace values, with other variables held at their averages
nd <- data.frame("terrace" = seq(min(sample$terrace), max(sample$terrace), length.out = 100),
                 "Z" = rep(mean(sample$Z), 100),
                 "Slope" = rep(mean(sample$Slope, na.rm = T), 100),
                 "Distance" = rep(mean(sample$Distance, na.rm = T), 100))

#use the glm model to predict forest loss values using these data
pred.vals <- predict(m1, newdata = nd, type = "response", se = T)

#plot relationship
plot(pred.vals$fit ~ nd$terrace, type = "l",
      xlab="Presence of Terraces",
      ylab="Likelihood of Forest Loss",
      xaxt = "n")
axis(1,at =c(0,1),
     labels = c(0,1))
```



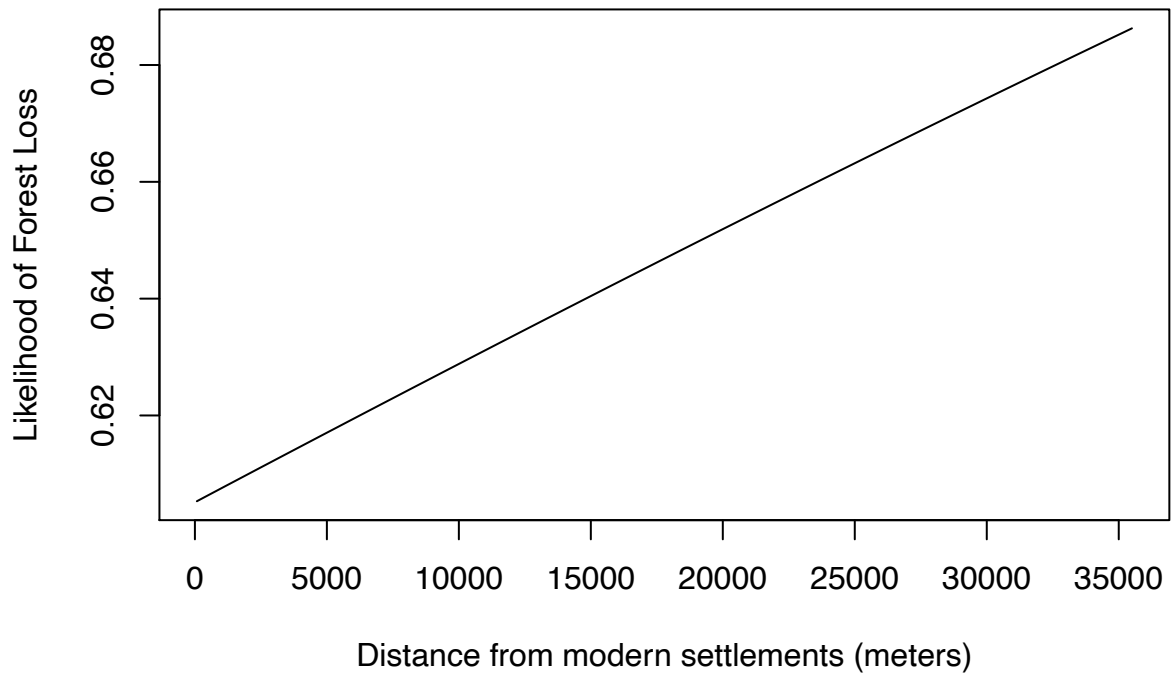

```
#relationship with elevation
nd2 <- data.frame("Z" = seq(min(sample$Z), max(sample$Z), length.out = 100),
                  "terrace" = rep(mean(sample$terrace), 100),
                  "Slope" = rep(mean(sample$Slope, na.rm = T), 100),
                  "Distance" = rep(mean(sample$Distance, na.rm = T), 100))

pred.vals2 <- predict(m1, newdata = nd2, type = "response", se = T)

plot(pred.vals2$fit ~ nd2$Z, type = "l",
     xlab="Elevation (masl)",
     ylab="Likelihood of Forest Loss")
```




```
#relationship with slope
nd3 <- data.frame("Slope" = seq(min(sample$Slope), max(sample$Slope), length.out = 100),
                  "terrace" = rep(mean(sample$terrace), 100),
                  "Z" = rep(mean(sample$Z, na.rm = T), 100),
                  "Distance" = rep(mean(sample$Distance, na.rm = T), 100))

pred.vals3 <- predict(m1, newdata = nd3, type = "response", se = T)

plot(pred.vals3$fit ~ nd3$Slope, type = "l",
     xlab="Slope (degrees)",
     ylab="Likelihood of Forest Loss")
```




```
#relationship with Distance
nd4 <- data.frame("Distance" = seq(min(sample$Distance), max(sample$Distance), length.out = 100),
                  "terrace" = rep(mean(sample$terrace), 100),
                  "Z" = rep(mean(sample$Z, na.rm = T), 100),
                  "Slope" = rep(mean(sample$Slope, na.rm = T), 100))

pred.vals4 <- predict(m1, newdata = nd4, type = "response", se = T)

plot(pred.vals4$fit ~ nd4$Distance, type = "l",
     xlab="Distance from modern settlements (meters)",
     ylab="Likelihood of Forest Loss")
```




```
sample$resid <- as.numeric(m1$residuals)
```

```
## pdf
##   2
```