# exploration

## Lyndsey Umsted

## 2022-07-12

```r
getwd()
```

```
## [1] "C:/Users/18586/Desktop/Summer Disease Internship/MacDonald-REU-Summer-22/models"
```

```r
aad <- read.csv("data/aad.csv")
aad <- subset(aad, !is.na(aad$Cutaneous.Leishmaniasis))
aad <-subset(aad, aad$Cutaneous.Leishmaniasis > 0)
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
mean_Cutaneous <- aad %>%
  group_by(Year) %>%
  summarise_at(vars(Cutaneous.Leishmaniasis), list(name = mean))
mean_Cutaneous
```

```
## # A tibble: 19 x 2
##     Year  name
##    <int> <dbl>
##  1  2001 0.959
##  2  2002 0.913
##  3  2003 0.945
##  4  2004 0.874
##  5  2005 0.802
##  6  2006 0.728
##  7  2007 0.735
##  8  2008 0.766
##  9  2009 0.928
## 10  2010 1.28
```

```
## 11   2011 1.05
## 12   2012 1.02
## 13   2013 1.05
## 14   2014 1.09
## 15   2015 1.00
## 16   2016 1.20
## 17   2017 0.870
## 18   2018 0.865
## 19   2019 1.69
```

```
plot(mean_Cutaneous, main = "Average Case Numbers between 2001 and 2019", xlab = "Year", ylab = "Cases

# mean_te_forest <- aad %>%
#   group_by(Year) %>%
#   summarise_at(vars(NDVI), list(name = mean))
# mean_te_forest
# plot(mean_te_forest)

fit <- lm(aad$Cutaneous.Leishmaniasis ~ aad$Year, data = aad)
summary(fit)
```

```
##
## Call:
## lm(formula = aad$Cutaneous.Leishmaniasis ~ aad$Year, data = aad)
##
## Residuals:
##     Min     1Q Median     3Q    Max
## -1.083 -0.868 -0.683 -0.091 90.787
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) -24.918605   4.390367  -5.676 1.39e-08 ***
## aad$Year      0.012881   0.002184   5.897 3.72e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.419 on 46295 degrees of freedom
## Multiple R-squared:  0.0007507,  Adjusted R-squared:  0.0007291
## F-statistic: 34.78 on 1 and 46295 DF,  p-value: 3.719e-09
```

```
abline(fit, col = "red")
```

## Average Case Numbers between 2001 and 2019
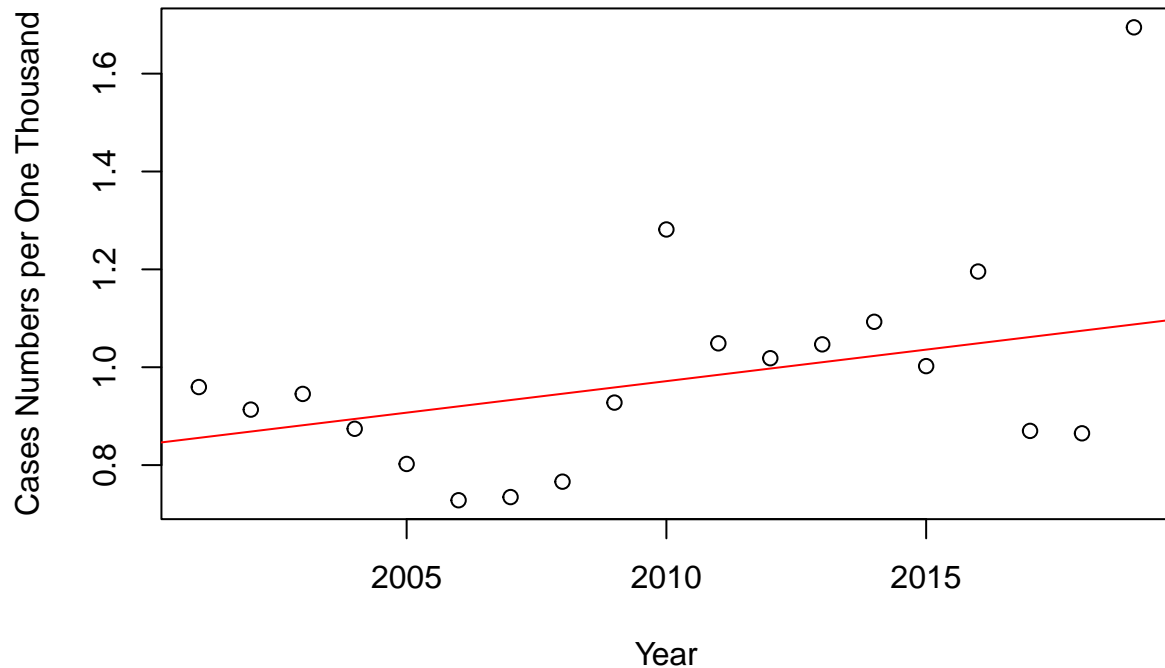


```
# set.seed(200)
# sample <- sample_n(aad, 500)
#
# fit <- lm(sample$Cutaneous.Leishmaniasis ~ sample$EVI, data = sample)
# summary(fit)
#
# confint(fit)
# plot(sample$EVI, sample$Cutaneous.Leishmaniasis)
# abline(fit, col = "red")
#
# library(MASS)
# boxcox(fit, lambda =seq(-0.1,0.1, 1), plotit = TRUE)
```

Diagnostic PlotS:

```
# par(mfrow = c(2,2), mgp = c(2,1,0), mar = c(3,3,3,1) + 0.1)
# plot(fitted(fit), residuals(fit), xlab = "fitted", ylab = "residuals")
# abline(h=0)
# title("Residuals vs Fitted")
# qqnorm(residuals(fit), ylab = "residuals", main = "QQ plot")
# qqline(residuals(fit))
# plot(sample$human_footprint, residuals(fit), xlab = "time", ylab = "resdisuals")
# abline(h=0)
# title("Residuals vs human_footprint")
```