R Project

Yuli Chen

October 26, 2018

```
# read in MozSyntheticMalaria.csv
mozdat <- read.csv(file="C:/Users/Yuli/OneDrive - The University of Colorado Denver/Do
cuments/Fall 2018/BIOS 6640 - Python & R/R/Data/MozSyntheticMalaria.csv",
           header=TRUE, sep=',')

mozdat2<-subset(mozdat, Epiyear < 2017)

# creating malaria incidence in case per 1000 population in children under 5
mozdat2$cases.u5 <- (mozdat2$malaria/(mozdat2$u5weight*mozdat2$Population_UN)*1000)

# total rain by District and epiyear
rainTot <- as.data.frame(tapply(mozdat2$rainTot, list(mozdat2$Province, mozdat2$Epiyea
r), sum))

# average temperature by District and epiyear
avgTemp <- as.data.frame(tapply(mozdat2$tavg, list(mozdat2$Province, mozdat2$Epiyea
r), mean))

# total under 5 cases per thousand by District and Epiyear
cpt <- as.data.frame(tapply(mozdat2$cases.u5, list(mozdat2$Province, mozdat2$Epiyea
r), sum))

# renaming column names
colnames(cpt) <- c("cpt10", "cpt11", "cpt12", "cpt13", "cpt14", "cpt15", "cpt16")
colnames(rainTot) <- c("rain10", "rain11", "rain12", "rain13", "rain14", "rain15", "ra
in16")
colnames(avgTemp) <- c("tavg10", "tavg11", "tavg12", "tavg13", "tavg14", "tavg15", "ta
vg16")

# combining total rainfall, average temperature, and total cases per 1000 by District
into a data frame
allStats <- as.data.frame(cbind(cpt, rainTot, avgTemp))

# take out Maputo City - duplicate data
allStats2<-allStats[-6,]
```

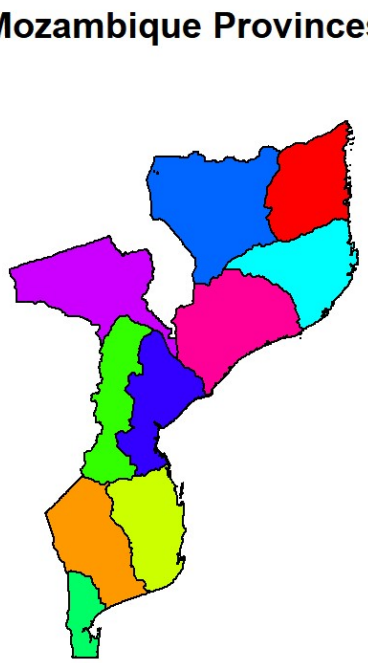
```
# reading in shapefile
poly1 <- readShapePoly("C:/Users/Yuli/OneDrive - The University of Colorado Denver/Doc
uments/Fall 2018/BIOS 6640 - Python & R/R/Data/Mozambique Admin1/mozambique_admin1.sh
p", IDvar = "NAME1")
```

```
## Warning: readShapePoly is deprecated; use rgdal::readOGR or sf::st_read
```

```
row.names(poly1)
```

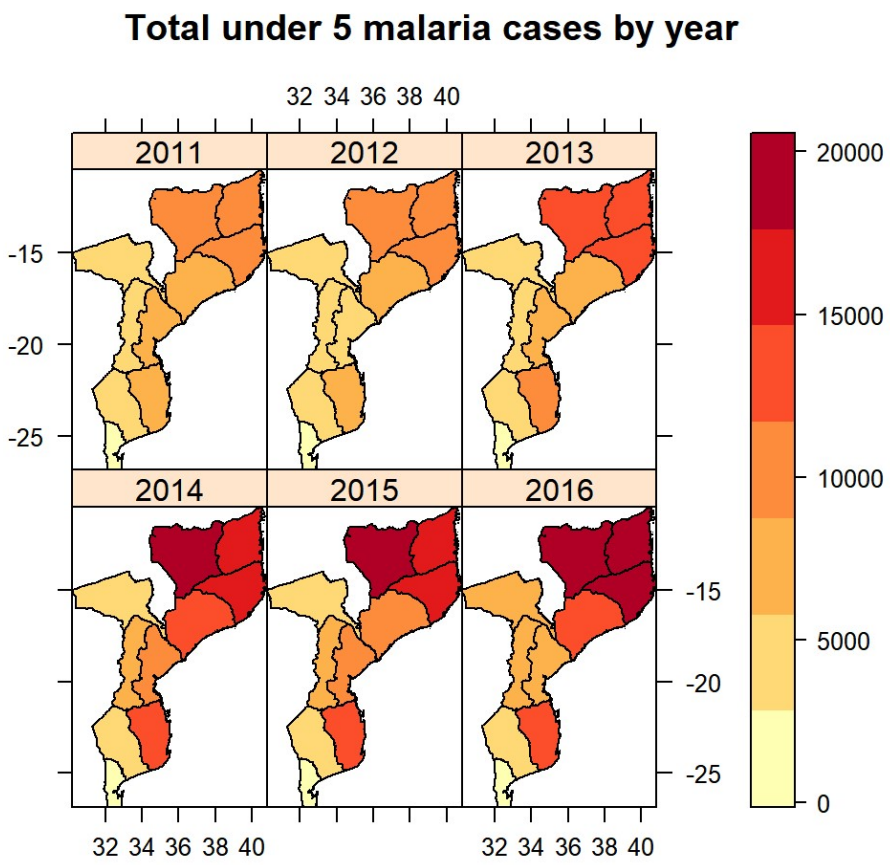
```
##  [1] "Cabo Delgado" "Gaza"         "Inhambane"    "Manica"       "Maputo"
##  [6] "Nampula"      "Nassa"        "Sofala"       "Tete"         "Zambezia"
```

```
# renaming rows to match row names in poly1
rownames(allStats2)<-c("Cabo Delgado", "Gaza", "Inhambane", "Manica", "Maputo", "Nampu
la", "Nassa", "Sofala", "Tete", "Zambezia")

# plotting the provinces of Mozambique
n<-length(poly1$NAME1)
plot(poly1, col=rainbow(n), main = 'Mozambique Provinces')
```

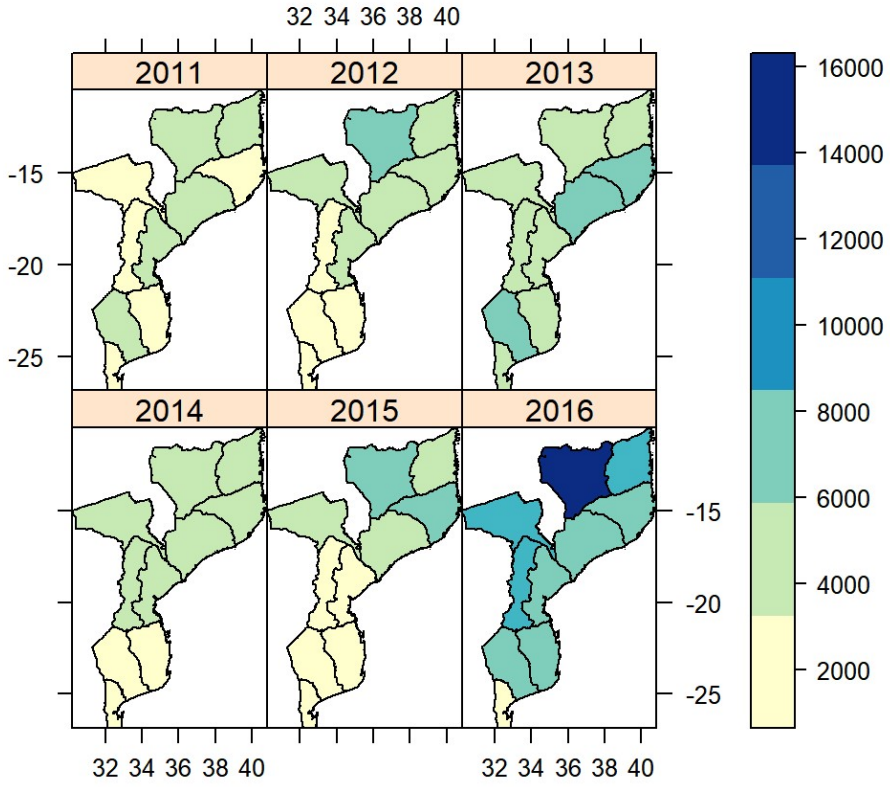


```
# combining data with shapefile
polydat <- SpatialPolygonsDataFrame(poly1, allStats2)

# load color palettes to use for mapping
tempPal <- brewer.pal(n = 7, name = "YlOrRd")
rainPal <- brewer.pal(n = 7, name = "YlGnBu")
```

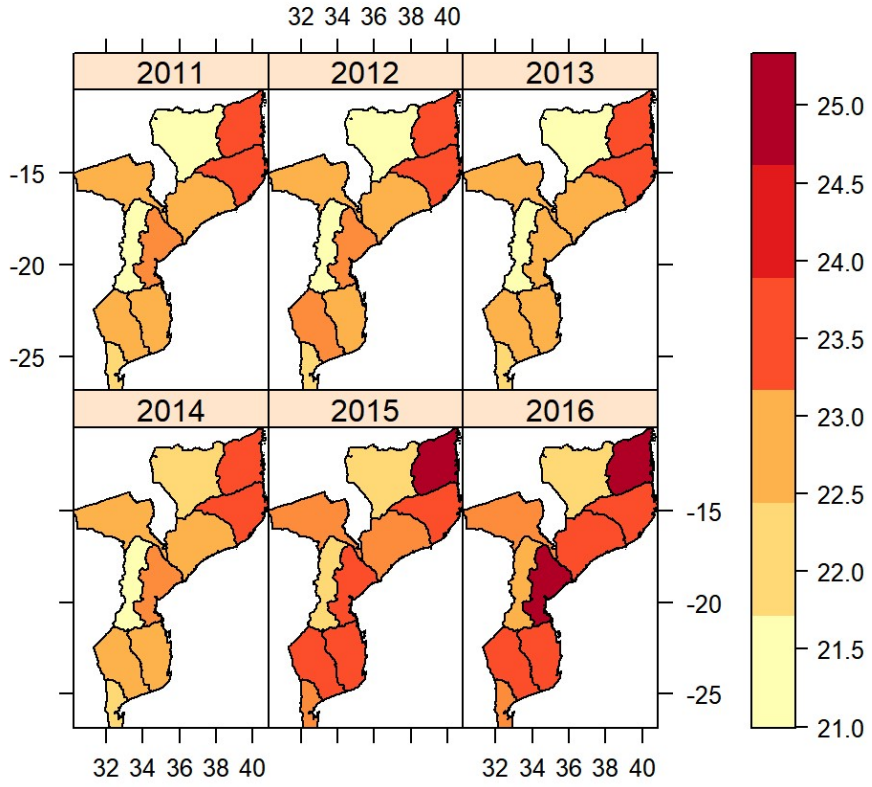
```
# map of total under 5 malaria cases per 1000 by province and year
spplot(polydat, c("cpt11", "cpt12", "cpt13", "cpt14", "cpt15", "cpt16"),
       names.attr = c("2011", "2012", "2013", "2014", "2015", "2016"),
       colorkey=list(space="right"), scales = list(draw = TRUE),
       main = "Total under 5 malaria cases by year",
       as.table = TRUE, col.regions = tempPal, col="black", cuts=6)
```




```
# map of total rainfall in Mozambique by province and year
spplot(polydat, c("rain11", "rain12", "rain13", "rain14", "rain15", "rain16"),
       names.attr = c("2011", "2012", "2013", "2014", "2015", "2016"),
       colorkey=list(space="right"), scales = list(draw = TRUE),
       main = "Total rainfall by year",
       as.table = TRUE, col.regions = rainPal, col="black", cuts=5)
```




```
# map of average temperature in Mozambique by province and year
spplot(polydat, c("tavg11", "tavg12", "tavg13", "tavg14", "tavg15", "tavg16"),
       names.attr = c("2011", "2012", "2013", "2014", "2015", "2016"),
       colorkey=list(space="right"), scales = list(draw = TRUE),
       main = "Average Temperature by Year",
       as.table = TRUE, col.regions = tempPal, col="black", cuts=5)
```




```
# combining rows so that each districts has 1 data point for each Epiyear
dat <- mozdat2 %>% group_by(Region, Province, Epiyear, District)%>%
  summarise(tavg_yr=mean(tavg), cases_u5=sum(cases.u5), totrain =sum(rainTot), tabove3
5=sum(tabove35), tbelow15=sum(tbelow15))
```

```
# creating column scatter plots
cases_plot<-ggplot(dat, aes(group=Epiyear, x=Epiyear, y=cases_u5))
rain_plot<-ggplot(dat, aes(group=Epiyear, x=Epiyear, y=totrain))
temp_plot<-ggplot(dat, aes(group=Epiyear, x=Epiyear, y=tavg_yr))

# putting plots together
require(gridExtra)
p1<-cases_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(widt
h=.2)) + xlab("") + ylab("Total Under 5 Cases per 1000")
p2<-rain_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(width
=.2)) + xlab("") + ylab ("Total Rainfall (mm)")
p3<-temp_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(width
=.2)) + xlab("Year") + ylab("Average Temperature (degrees Celcius)")
grid.arrange(p1,p2,p3, ncol=1)
```


```
# creating plots for extreme temperatures
thigh_plot<-ggplot(dat, aes(group=Epiyear, x=Epiyear, y=tabove35))
tlow_plot<-ggplot(dat, aes(group=Epiyear, x=Epiyear, y=tbelow15))

# putting plots together - extreme temperatures
require(gridExtra)
plot1<-cases_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(w
idth=.2)) + xlab("") + ylab("Under 5 Cases per 1000")
plot2<-temp_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(wi
dth=.2)) + xlab("") + ylab("Mean Temperature (Celcius)")
plot3<-thigh_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(w
idth=.2)) + xlab("") + ylab("Days Above 35 degrees C")
plot4<-tlow_plot+geom_jitter(alpha=0.5, aes(color=Region), position=position_jitter(wi
dth=.2)) + xlab("Year") + ylab("Days Below 15 degrees C")
grid.arrange(plot1,plot2,plot3,plot4, ncol=1)
```


```
# creating lagged variables
mozdat3 <- mozdat2 %>%
  group_by(DISTCODE) %>%            # creating lagged total rainfall variable
  mutate(rainTot2 = lag(rainTot, 2), # 2 week lag
         rainTot4 = lag(rainTot, 4), # 4 week lag
         rainTot8 = lag(rainTot, 8), # 8 week lag
         tavg2= lag(tavg, 2), # creating lagged weely average temperatures
         tavg4 = lag(tavg, 4),
         tavg8 = lag(tavg, 8))
```

```
# lagged plot: plotting under 5 incidence and rainfall
ggplot(data = mozdat3) +
  geom_smooth(mapping = aes(x = Epiweek, y = cases.u5,  color= "Under 5 Malaria Case
s")) +
  geom_smooth(mapping = aes(x = Epiweek, y = rainTot, color= "Total Rainfall(mm)")) +
  geom_smooth(mapping = aes(x = Epiweek, y = rainTot2, color= "Rainfall, Lagged 2 week
s")) +
  geom_smooth(mapping = aes(x = Epiweek, y = rainTot4, color= "Rainfall, Lagged 4 week
s")) +
  geom_smooth(mapping = aes(x = Epiweek, y = rainTot8, color= "Rainfall, Lagged 8 week
s")) +
  facet_wrap(~ Region, nrow=2) +
  scale_y_continuous(sec.axis = sec_axis(~.*2, name = "Total Weekly rainfall (mm)")) +
  labs(x = "Epidemiology week", y = "Cases per 1,000")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

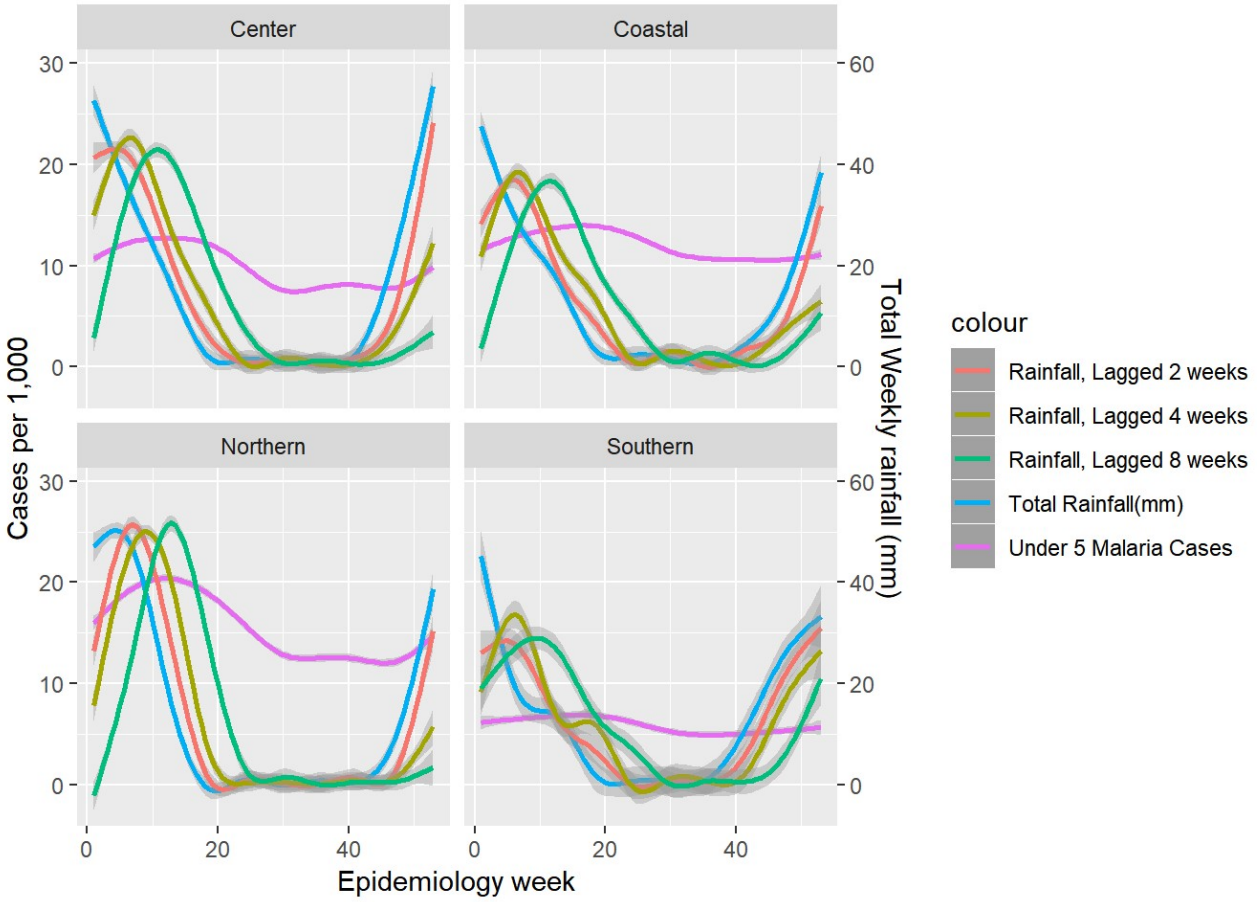
```
## Warning: Removed 284 rows containing non-finite values (stat_smooth).
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 568 rows containing non-finite values (stat_smooth).
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 1136 rows containing non-finite values (stat_smooth).
```




```
# lagged plot: plotting under 5 incidence and average temperature
ggplot(data = mozdat3) +
  geom_smooth(mapping = aes(x = Epiweek, y = cases.u5,  color= "Under 5 Malaria Case
s")) +
  geom_smooth(mapping = aes(x = Epiweek, y = tavg, color= "Average Temperature")) +
  geom_smooth(mapping = aes(x = Epiweek, y = tavg2, color= "Average Temperature, Lagge
d 2 weeks")) +
  geom_smooth(mapping = aes(x = Epiweek, y = tavg4, color= "Average Temperature, Lagge
d 4 weeks")) +
  geom_smooth(mapping = aes(x = Epiweek, y = tavg8, color= "Average Temperature, Lagge
d 8 weeks")) +
  facet_wrap(~ Region, nrow=2) +
  scale_y_continuous(sec.axis = sec_axis(~.*2, name = "Average Temperature")) +
  labs(x = "Epidemiology week", y = "Cases per 1,000")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 284 rows containing non-finite values (stat_smooth).
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 568 rows containing non-finite values (stat_smooth).
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 1136 rows containing non-finite values (stat_smooth).
```