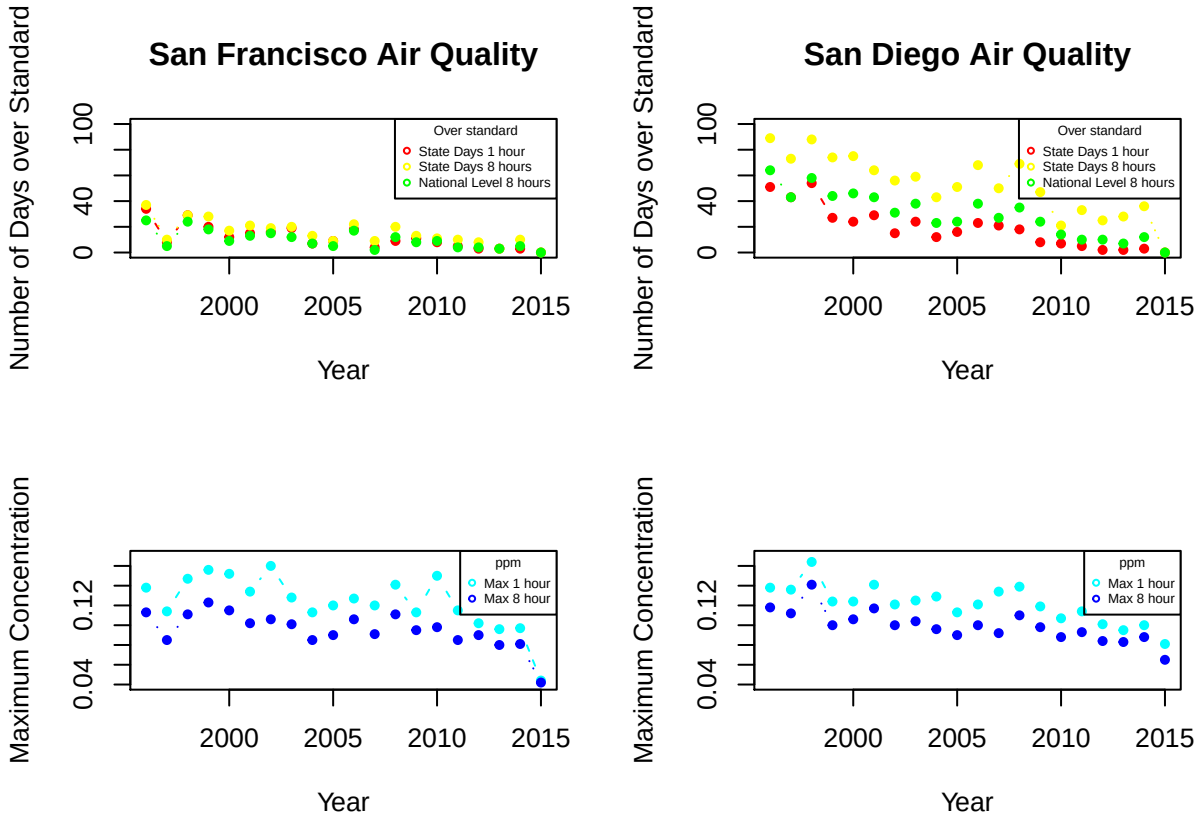
Assignment2_Xin Huang

Answer for Question 1

Run the following code to generate the figures:

```
#set workspace and read file
setwd("/Users/xinhuang/Google Drive/CSC 522 R Language Programming /Homework2")
sfData <- read.csv("AirQualitySanFranciscoData.csv", header = TRUE)
sdData <- read.csv("AirQualitySanDiegoData.csv", header = TRUE)

# set layout
mat <- matrix(1 : 4, nrow = 2)
layout(mat)

with(sfData, {
    plot(Year, ExceedStateLevelDays1.hr, col = "red", ylim = c(0, 100),
         type = "b", ylab = "Number of Days over Standard",
         pch = 20, lty = 2)
    points(Year, ExceedStateLevelDays8.hr, col = "yellow",
           type = "b", pch = 20, lty = 3)
    points(Year, ExceedNationalLevelDays8.hr, col = "green",
           type = "b", pch = 20, lty = 4)
    title(main = "San Francisco Air Quality")
    legend("topright",
           c("State Days 1 hour", "State Days 8 hours", "National Level 8 hours"),
           col = c("red", "yellow", "green"),
           pch = 1, cex = 0.5, title = "Over standard")
})

with(sfData, {
    plot(Year, MaxConcentration1.hr, col = "cyan", ylim = c(0.04, 0.17),
         type = "b", ylab = "Maximum Concentration", pch = 20, lty = 2)
    points(Year, MaxConcentration8.hr, col = "blue",
           type = "b", pch = 20, lty = 3)
    legend("topright", c("Max 1 hour", "Max 8 hour"),
           col = c("cyan", "blue"), pch = 1, cex = 0.5, title = "ppm")
})

with(sdData, {
    plot(Year, ExceedStateLevelDays1.hr, col = "red", ylim = c(0, 100),
         type = "b", ylab = "Number of Days over Standard",
         pch = 20, lty = 2)
    points(Year, ExceedStateLevelDays8.hr, col = "yellow",
           type = "b", pch = 20, lty = 3)
    points(Year, ExceedNationalLevelDays8.hr, col = "green",
           type = "b", pch = 20, lty = 4)
    title(main = "San Diego Air Quality")
    legend("topright",
           c("State Days 1 hour", "State Days 8 hours", "National Level 8 hours"),
           col = c("red", "yellow", "green"),
           pch = 1, cex = 0.5, title = "Over standard")
```

```
})

with(sdData, {
    plot(Year, MaxConcentration1.hr, col = "cyan", ylim = c(0.04, 0.17),
         type = "b", ylab = "Maximum Concentration",
         pch = 20, lty = 2)
    points(Year, MaxConcentration8.hr, col = "blue",
           type = "b", pch = 20, lty = 3)
    legend("topright", c("Max 1 hour", "Max 8 hour"),
           col = c("cyan", "blue"), pch = 1, cex = 0.5, title = "ppm")
})

mtext("Main title", line=2, font=2, cex=1.2, outer = TRUE)
```

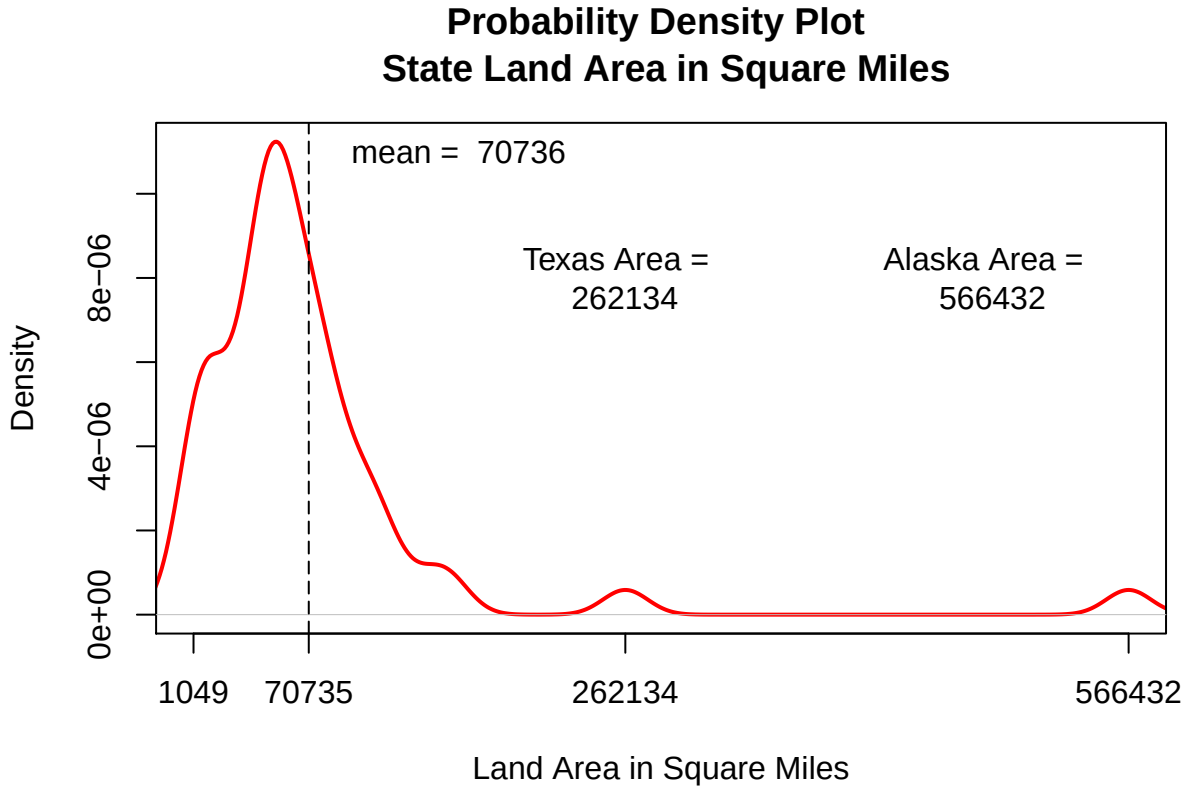


Answer for Question 2

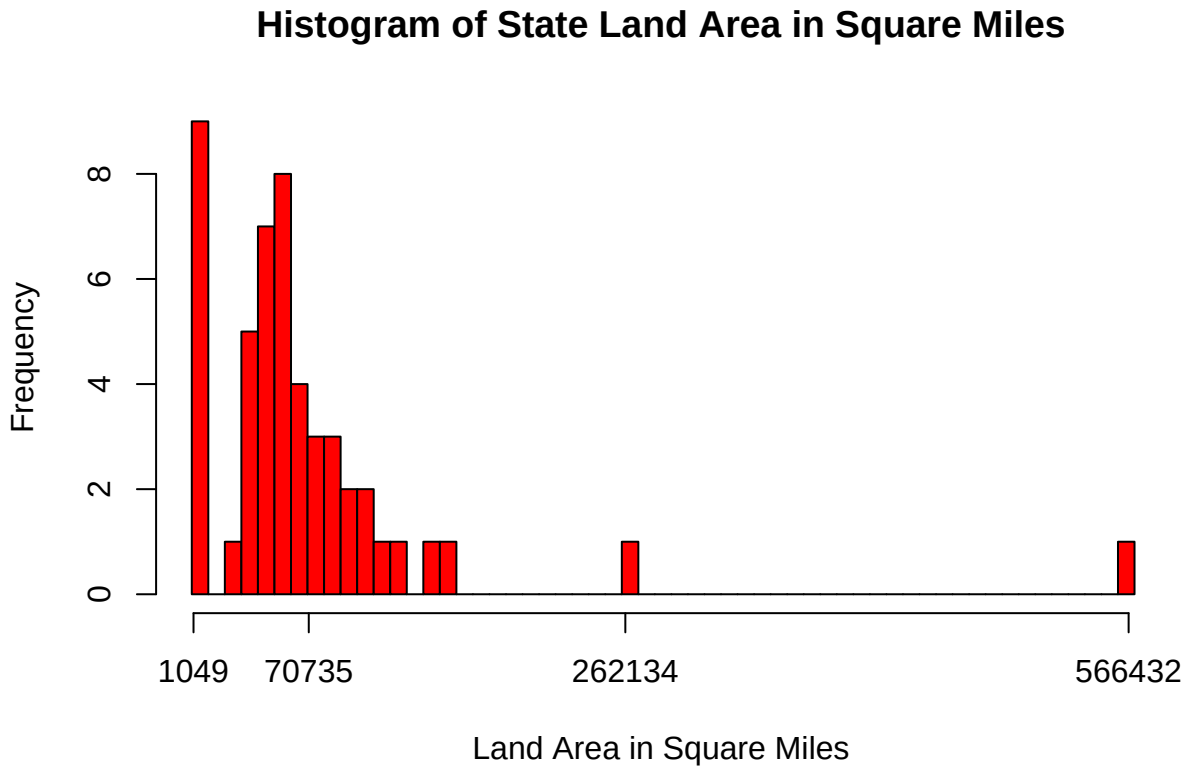
Run the following code to generate the figures:

```
#Question 2
myState <- as.data.frame(state.x77)
d <- density(myState$Area)
axisV <- as.integer(c(min(myState$Area),
                      mean(myState$Area),
                      myState[rownames(myState)== "Texas",
                              colnames(myState)=="Area"],
                      max(myState$Area)))
```

```
plot(d, col = "red", lwd = 2,
    xlim=c(min(myState$Area),max(myState$Area)),
    xlab = "Land Area in Square Miles",
    ylab = "Density",
    xaxt = "n",
    main = "Probability Density Plot \n State Land Area in Square Miles")
axis(1, at = axisV, labels = axisV)
abline(v=mean(myState$Area),col = "black", lty = 5)
text(2 * mean(myState$Area) + 20000, 11e-06,
     paste("mean = ",round(mean(myState$Area), digits=0)))
text(myState$Area[rownames(myState)=="Alaska"]-85000, 8e-06,
     paste("Alaska Area = \n",myState$Area[rownames(myState)=="Alaska"]))
text(myState$Area[rownames(myState)=="Texas"]-3000, 8e-06,
     paste("Texas Area = \n",myState$Area[rownames(myState)=="Texas"]))
```

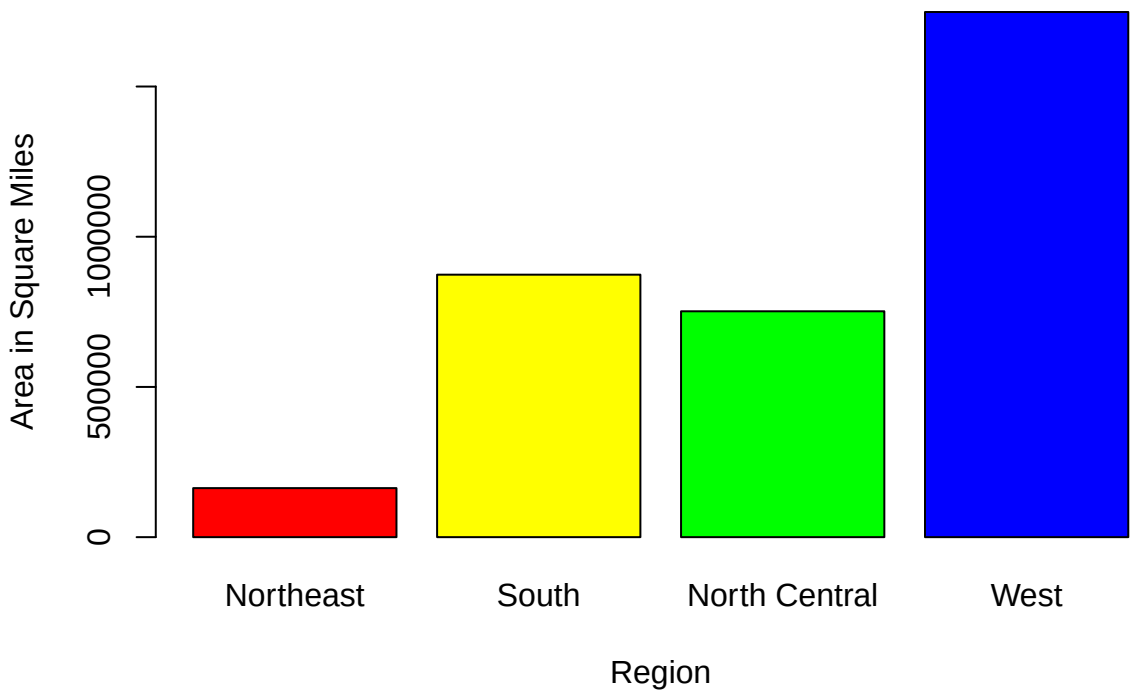



```
hist(myState$Area,
     breaks= 50,
     col="red",
     xaxt = "n",
     xlim=c(min(myState$Area),max(myState$Area)),
     xlab = "Land Area in Square Miles",
     main = "Histogram of State Land Area in Square Miles")
axis(1, at = axisV, labels = axisV)
```



Answer for Question 3

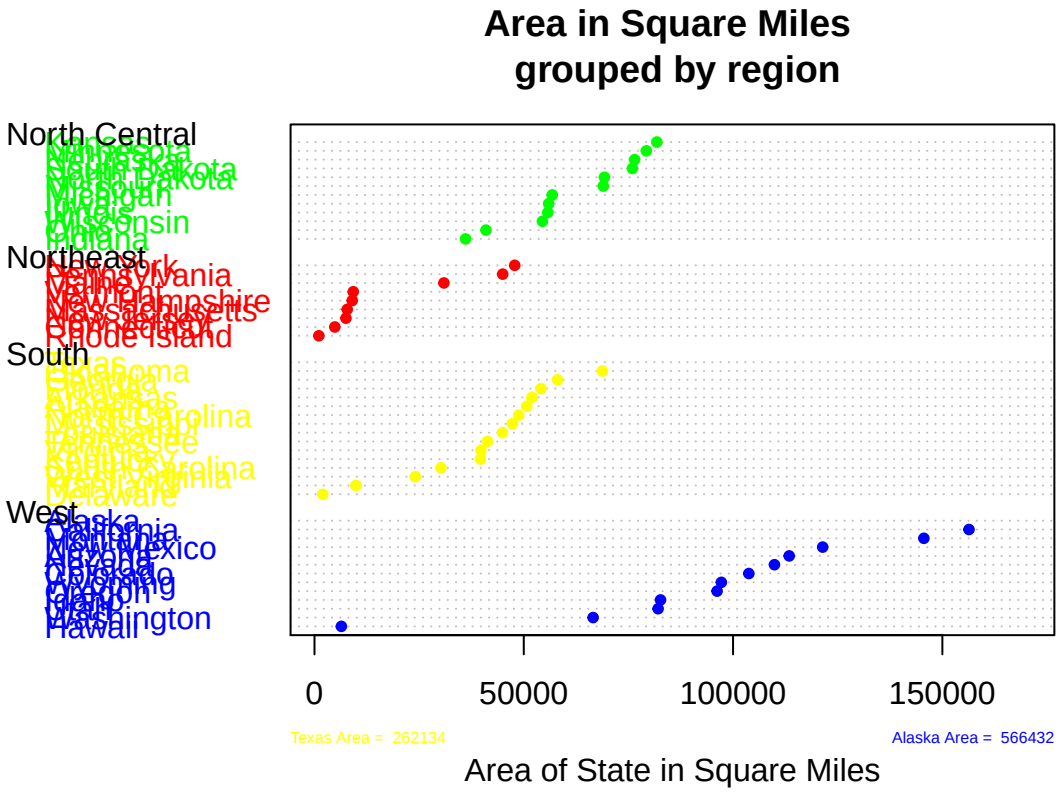
```
#Question 3
##Figure 1
myState <-as.data.frame(cbind(state.x77, region = state.region))
myState <- cbind(myState, regionName = levels(state.region)[state.region])

stateSum <- myState %>%
    group_by(regionName) %>%
    summarise_all(sum)
target <- c("Northeast", "South", "North Central", "West")
stateSum <- stateSum[match(target, stateSum$regionName), ]

barplot(stateSum$Area,
        names.arg = stateSum$regionName,
        col = c("red", "yellow", "green", "blue"),
        xlab = "Region",
        ylab = "Area in Square Miles",
        ylim = c(0, 1700000),
        main = "Area By Region")
```



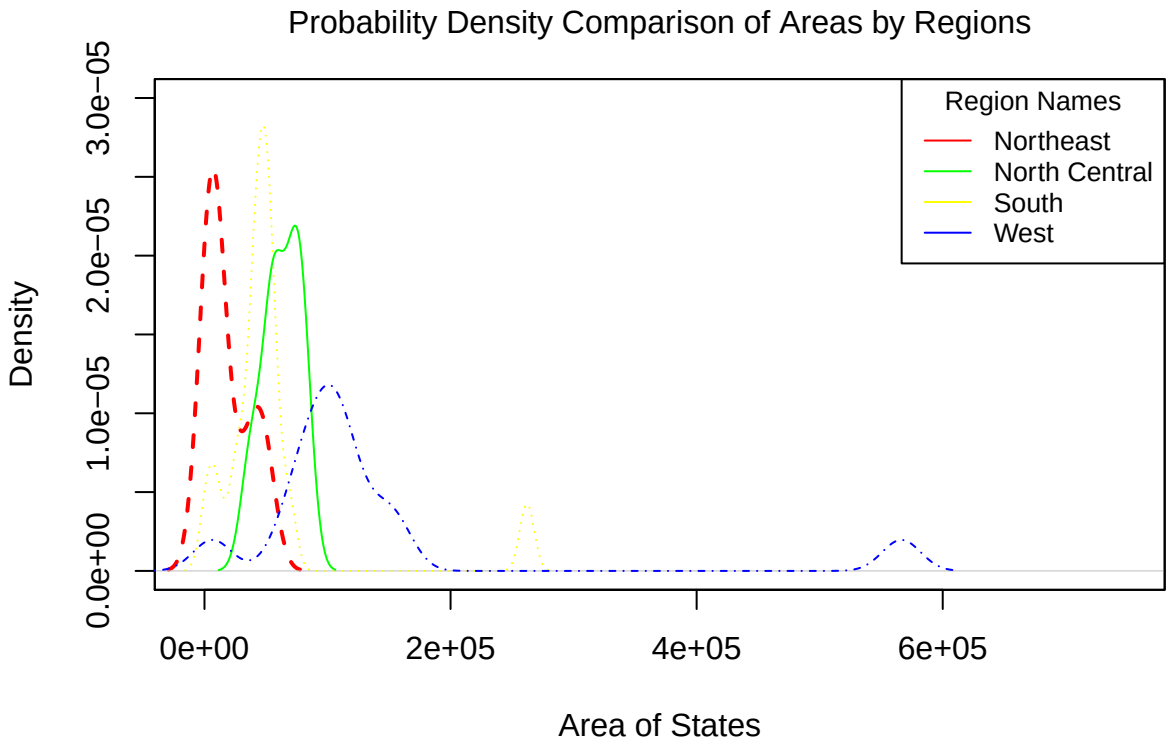

```
##Figure 2
myState <- myState[order(myState$regionName, myState$Area),]
myState$regionName <- factor(myState$regionName)
myState$color[myState$regionName == "Northeast"] <- "red"
myState$color[myState$regionName == "South"] <- "yellow"
myState$color[myState$regionName == "North Central"] <- "green"
myState$color[myState$regionName == "West"] <- "blue"

myState$Area[rownames(myState)=="Texas"]
```

```
## [1] 262134
```

```
TexasArea <- myState$Area[rownames(myState)=="Texas"]
AlaskaArea <- myState$Area[rownames(myState)=="Alaska"]

dotchart(myState$Area,
         labels= rownames(myState),
         cex= 1,
         groups = myState$regionName,
         gcolor = "black",
         color = myState$color,
         xlim = c(min(myState$Area), 170000),
         main = "Area in Square Miles \n grouped by region",
         xlab ="Area of State in Square Miles",
         pch = 20)
mtext(paste("Texas Area = ", TexasArea), side = 1, line= 2, adj = 0, col = "yellow", cex = .5)
mtext(paste("Alaska Area = ", AlaskaArea), side = 1, line= 2, adj = 1, col = "blue", cex = .5)
```



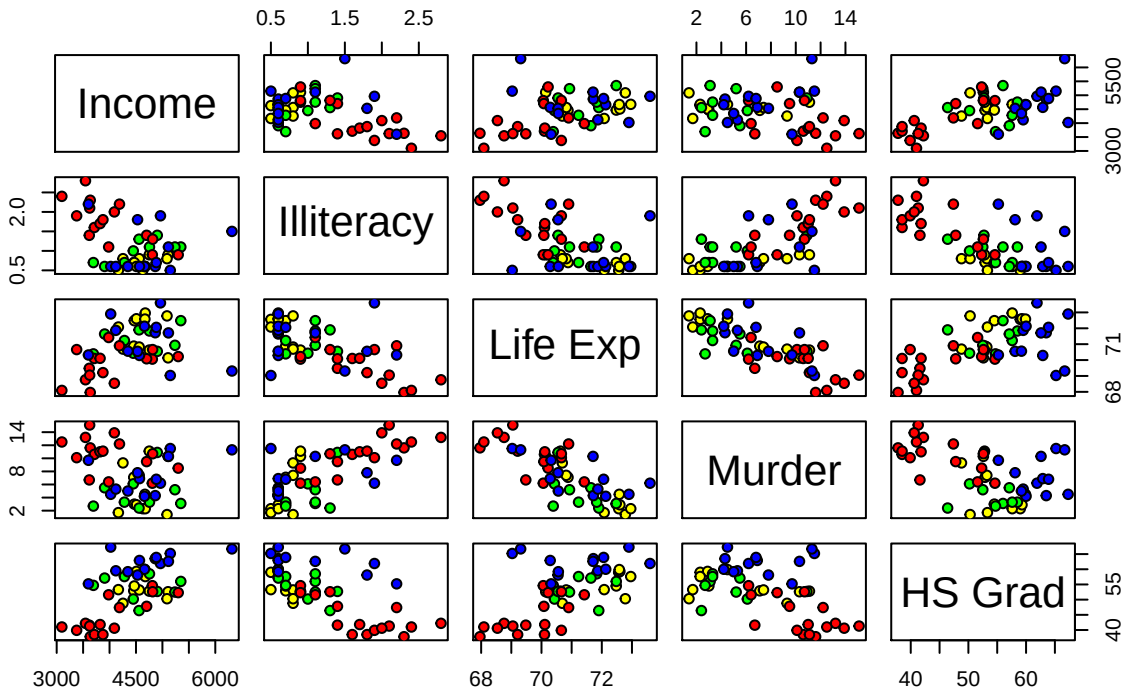

```
#Figure 3
d1 <- density(myState[which(myState$regionName == "Northeast"),]$Area)
d2 <- density(myState[which(myState$regionName == "South"),]$Area)
d3 <- density(myState[which(myState$regionName == "North Central"),]$Area)
d4 <- density(myState[which(myState$regionName == "West"),]$Area)

plot(d1, col = "red", lwd = 2,
     xlim = c(-10000, 750000),
     xlab = "Area of States",
     ylim = c(0, 0.000030),
     ylab = "Density",
     lty = 2, main = "")
box(lty = 1, col = 'black')
lines(d2, col = "yellow", lty = 3)
lines(d3, col = "green", lty = 1)
lines(d4, col = "blue", lty = 4)
legend("topright", c("Northeast", "North Central", "South", "West"),
       col = c("red", "green", "yellow", "blue"),
       lty = 1, cex = 0.8,
       title = "Region Names")
mtext("Probability Density Comparison of Areas by Regions", line = 1, outer = FALSE)
```



Answer for Question 4

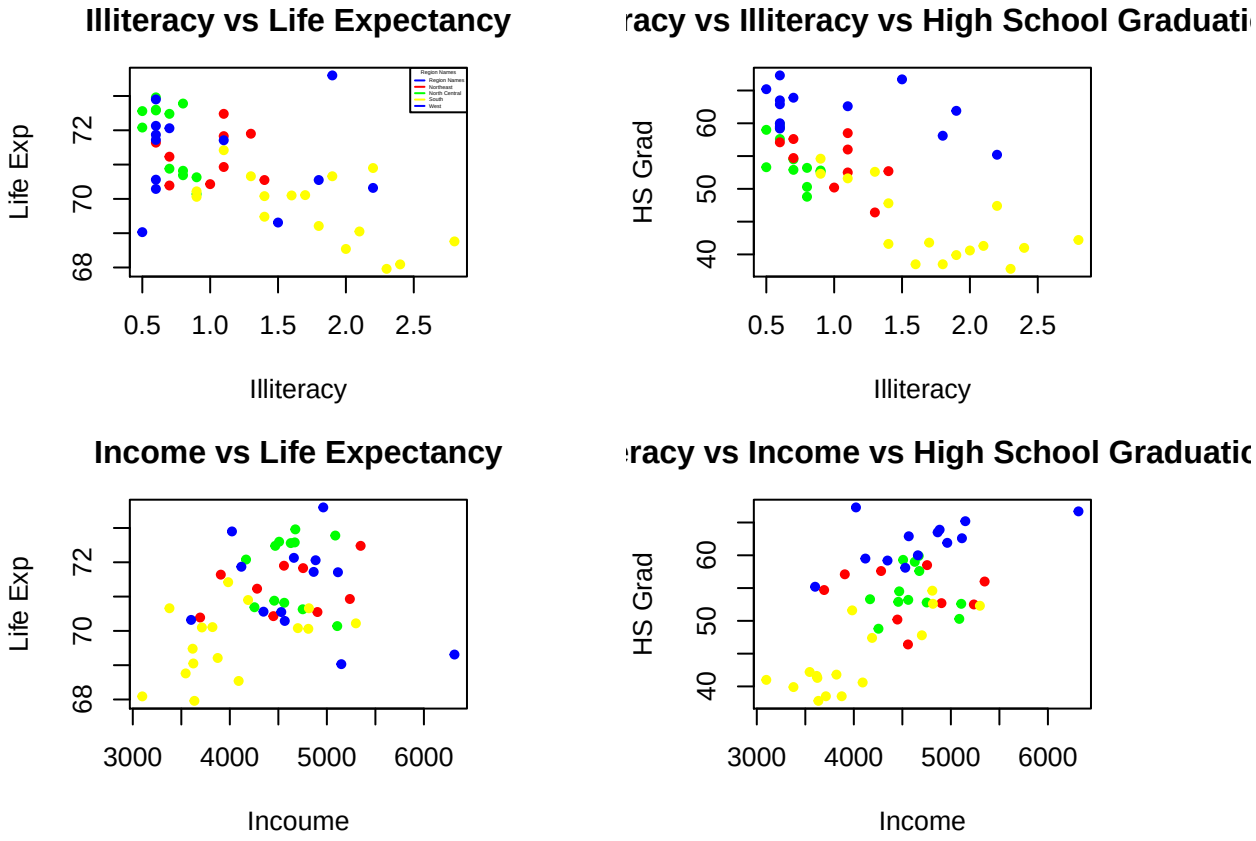
Run the following code to generate the figures:

```
##Question 4
myState <-as.data.frame(cbind(state.x77, region = state.region))
myState <- cbind(myState, regionName = levels(state.region)[state.region])
myState <- myState[order(myState$regionName, myState$Area),]
myState$regionName <- factor(myState$regionName)
myState$color[myState$regionName == "Northeast"] <- "red"
myState$color[myState$regionName == "South"] <- "yellow"
myState$color[myState$regionName == "North Central"] <- "green"
myState$color[myState$regionName == "West"] <- "blue"
##Figure 1
par(mar = c(4, 4, 3, 5), oma = c(0, 0, 0, 0))
pairs(myState[,2:6], main="Pairs Plot", pch = 21,
      bg = c("green", "red", "yellow", "blue")[unclass(myState$region)])
```




```
##Figure 2
mat <- matrix(1 : 4, nrow = 2)
layout(mat)

plot(myState$Illiteracy,  myState$`Life Exp`,
     col = myState$color, pch = 20, bg = myState$color,
     xlab = "Illiteracy",
     ylab = "Life Exp",
     main = "Illiteracy vs Life Expectancy")

legend("topright",
       legend = c("Region Names", "Northeast", "North Central", "South", "West"),
       col = c("blue","red", "green", "yellow", "blue"),
       lty = 1, cex = 0.2,
       title = "Region Names")

plot(myState$Income,  myState$`Life Exp`,
     col = myState$color, pch = 20, bg = myState$color,
     xlab = "Incoume",
     ylab = "Life Exp",
     main = "Income vs Life Expectancy")

plot(myState$Illiteracy,  myState$`HS Grad`,
     col = myState$color, pch = 20, bg = myState$color,
     xlab = "Illiteracy",
     ylab = "HS Grad",
     main = "Illiteracy vs Illiteracy vs High School Graduation")

plot(myState$Income,  myState$`HS Grad`,
     col = myState$color, pch = 20, bg = myState$color,
```

```
      xlab = "Income",
      ylab = "HS Grad",
      main = "Illiteracy vs Income vs High School Graduation")
```

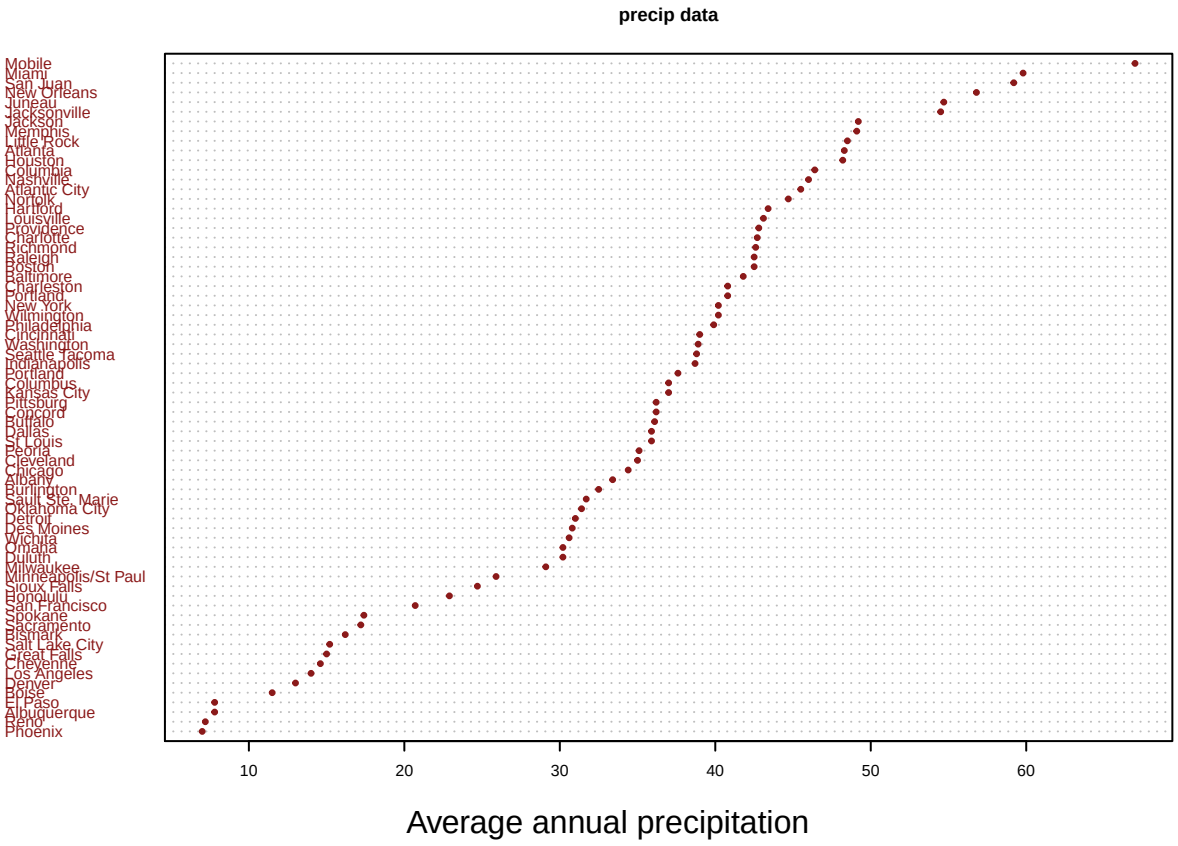


You can drive the following observations:
1. The higher education one has, the higher income he might have
2. The lower percentage of one region, the longer time spane of that region can have

Answer for Question 5

Using precip dataset

```
## dotchart
require(graphics)
dotchart(precip[order(precip)], main = "precip data",
        pch = 20, cex = 0.5, color = "firebrick4")
title(sub = "Average annual precipitation")
```



Using state.x77 dataset

Use *plot3d*() to plot column $Income, Illiteracy, LifeExp$

```
myState <-as.data.frame(cbind(state.x77, region = state.region))
attach(myState)
names(myState)
```

```
## [1] "Population" "Income"     "Illiteracy" "Life Exp"   "Murder"
## [6] "HS Grad"    "Frost"      "Area"       "region"
```

```
plot3d(Income , Illiteracy, myState$`Life Exp`, col="red", size=10)
```