**Question 1**

This question uses the lattice package. Show the code you used to generate the graphs. Use the state.region and state.x77 data set (recall typing ?state.x77 into the console will provide a discription of this data set).

**Question 1a**

Investigate the population density of USA states within the four regions: Northeast, South, North Central and West. Observe the graph of State Population as a Function of Area. Create Figure 1.

**Answer 1a**

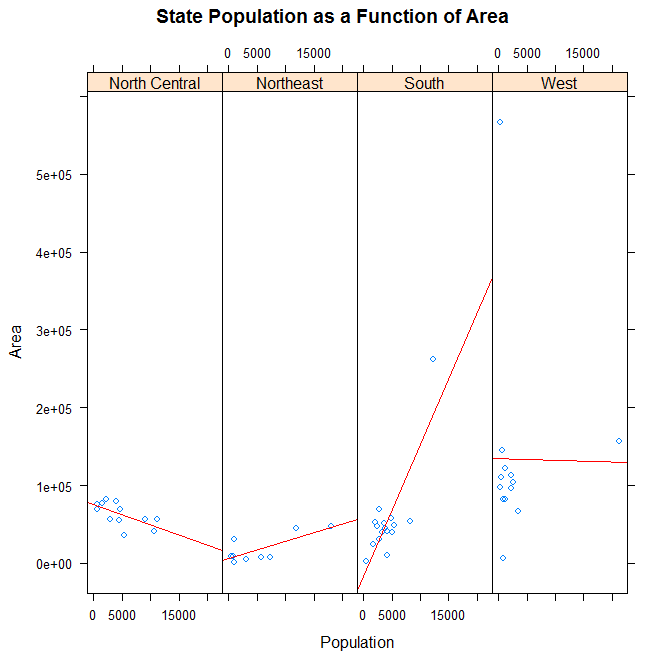
> usaState <-as.data.frame(cbind(state.x77,region=state.region))

> usaState <- cbind(usaState,regionName=levels(state.region)[state.region])

> head(usaState)

> library(lattice)

>  xyplot(usaState$Area ~ usaState$Population   | as.factor(usaState$regionName), layout = c(4, 1),  
       main = "State Population as a Function of Area", xlab="Population",ylab="Area",  
       panel = function(x, y, ...) {  
  panel.xyplot(x, y, ...)   
  panel.lmline(x, y, col = 2)   
  })



**Question 1C**

Use the cloud() function from the lattice package to create the 3D plot of States (Figure 2). Can you make this plot look more informative or interesting?

**Answer 1C**

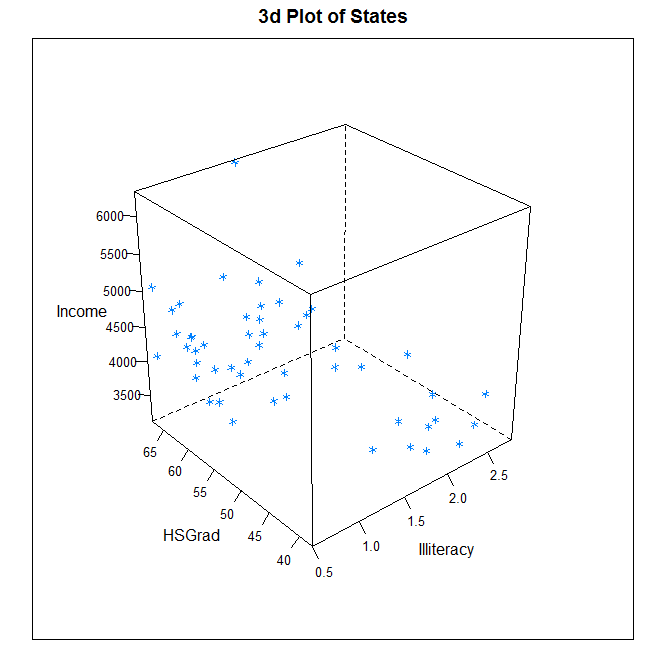
> state.x77.df <- data.frame(state.x77)

> stateName(state.x77.df) <- c("Population", "Income", "Illiteracy", "LifeExp", "Murder", "HSGrad", "Frost", "Area")

> attach(state.x77.df)

> library(lattice)

> cloud(Income ~ Illiteracy \* HSGrad, xlim=range(Illiteracy), ylim=range(HSGrad), zlim=range(Income), main="3d Plot of States", scales = list(distance = rep(1, 3), arrows = FALSE))

****

**Question 1d**

For this question the HS Grad column has been divided into three groups. For the first group, the HS Grad percentage is less than 50. In the second group the HS Grad percentage is between 50 and 57, while the HS Grad percentage for the third group is greater than 57. Use the lattice coplot() function to create Figure 3 .

**Answer 1 d**

> state.x77.df <- data.frame(state.x77)

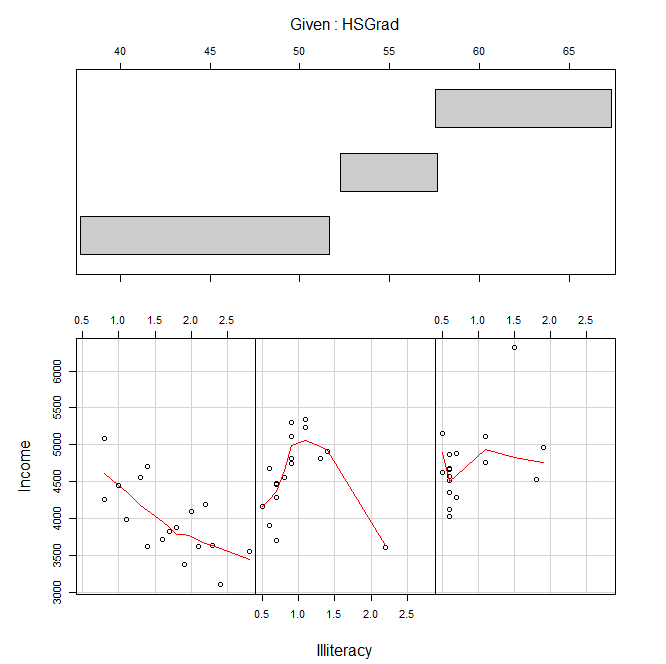
> stateName(state.x77.df) <- c("Population", "Income", "Illiteracy", "LifeExp", "Murder", "HSGrad", "Frost", "Area")

> attach(state.x77.df)

> m.interval = co.intervals(state.x77.df$HSGrad, number = 3, overlap = 0)

> coplot(Income~Illiteracy|HSGrad, data = state.x77.df, given.values = m.interval,rows=1,

+ panel = function(x, y, ...) panel.smooth(x, y, span = 0.7, ...))

****

**Question 2**

Create the graphs for this question using the ggplot2 package.

**Question 2a**

Recreate the graph of State Population as a Function of Area from Question 1a using the qplot() function from the ggplot2 package (Figure 4). Can you make the Population Axis look better?

**Answer 2a**

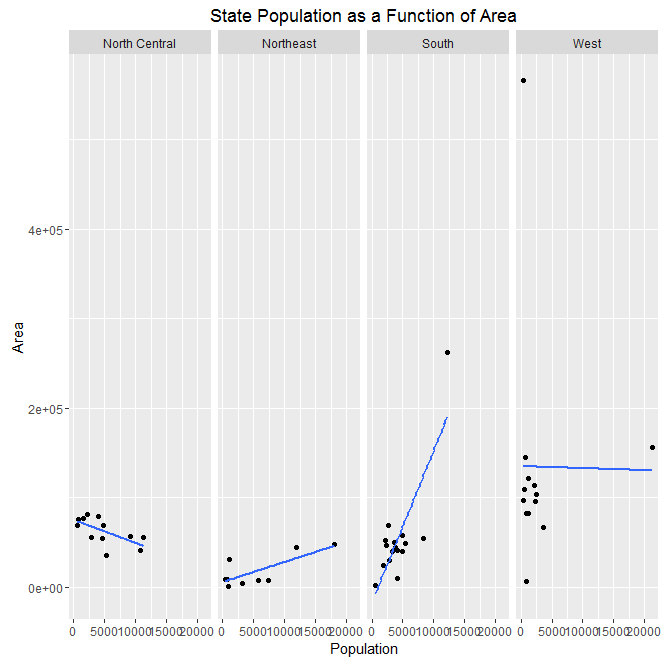
> rm(list=ls())

> library(ggplot2)

> statesPop <-as.data.frame(cbind(state.x77,region=state.region))

> statesPop <- cbind(statesPop,regionName=levels(state.region)[state.region])

> qplot(Population, Area, data=states, facets= . ~ regionName, main = "State Population as a Function of Area") +geom\_smooth(method = "lm", se = FALSE)



**Question 2b**

Investigate the land areas of USA regions. State regions can be found in the state.region data set. Use the qplot function from the ggplot2 package. Show the code you used to generate the graphs. Use the data sets state.x77 and state.region. Notice the axis for the charts. Use the grid.arrange() function from the package gridExtra to put two plots on the same page. Create the graphs and charts exactly as in Figure 5

**Answer 2b**

> library(gridExtra)

> Illiteracy <- as.numeric(as.character(states$Illiteracy))

> LifeExpectancy<- as.numeric(as.character(states$`Life Exp`))

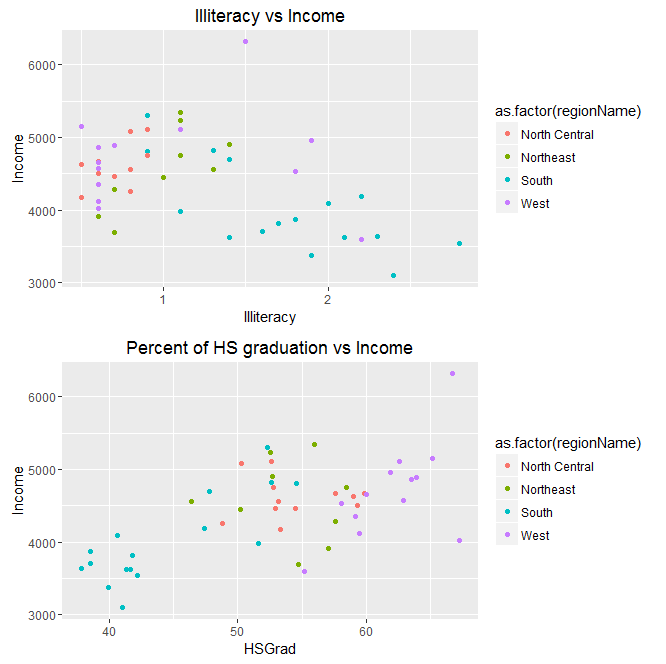
> HSGrad <- as.numeric(as.character(states$`HS Grad`))

> Income <- as.numeric(as.character(states$Income))

> g1<- qplot(Illiteracy, Income, data=states, col=as.factor(regionName), main="Illiteracy vs Income")

> g2<- qplot(HSGrad, Income, data=states, col=as.factor(regionName), main="Percent of HS graduation vs Income")

> grid.arrange(g1, g2, ncol=1)



**Question 3**

The figure for this question is a replica of one from Wickham’s ggplot2 text. To produce the figure, use the following data frame: > df<- data.frame( + x=c(3,1,5), + y=c(2,4,6), + label=c("a","b","c") + ) Next create a variable named myPlot as follows: > myPlot<-ggplot(df,aes(x,y,label=label))+ + xlab(NULL)+ylab(NULL) Now create eight graphs. Each of these starts with myPlot and adds only one geom layer along with specifying a title. The first graph adds a geom point() layer and annotates with the title ggtitle(”geom point”). Recall that you would need to add an extra line of code (ie. add the line pP1) to get the plot to print. > pP1<-myPlot+geom\_point()+ggtitle("geom\_point") Each of the remaining seven graphs need to have one of the following layers added: geom bar(stat=”identity”), geom line(), geom area(), geom path(), geom text(),geom tile(), geom polygon(). Exactly replicate the graph as shown in Figure 6. Show your code and explain the charts.

**Answer 3**

> df<- data.frame(x=c(3,1,5),y=c(2,4,6),label=c("a","b","c"))

> myPlot<-ggplot(df,aes(x,y,label=label)) + xlab(NULL)+ylab(NULL)

> a<-myPlot+geom\_point()+ggtitle("geom\_point")

> b<-myPlot+geom\_bar(stat="identity")+ggtitle("geom\_bar")

> c<-myPlot+geom\_line()+ggtitle("geom\_line")

> d<-myPlot+ geom\_area()+ggtitle("geom\_area")

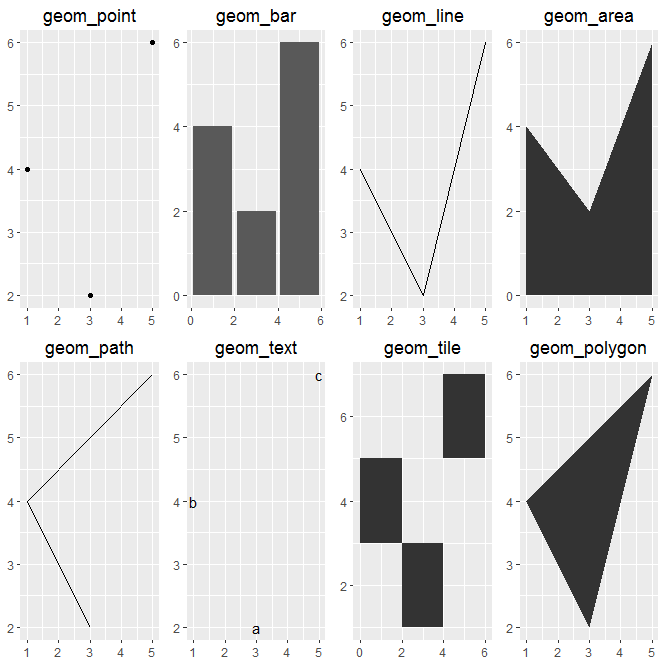
> e<-myPlot+ geom\_path()+ggtitle("geom\_path")

> f<-myPlot+ geom\_text()+ggtitle("geom\_text")

> g<-myPlot+geom\_tile()+ggtitle("geom\_tile")

> h<-myPlot+ geom\_polygon()+ggtitle("geom\_polygon")

> grid.arrange(a, b,c,d,e,f,g,h, ncol=4)



**Question 5**

Use the ggplot2 package. Show the code you used to generate the graphs. What about coloring a USA map using either the ”state.division”or ”state.region” the ”state” data set? What about making a map of another country? Be creative. Generate some graphs and charts using a data set of your choice.

**Answer 5**

> library(ggplot2)

> library(plyr)

> library(maps)

> library(Hmisc)

> library(stringi)

> data(state)

> states <- map\_data("state")

> colnames(states)[5] <- "State"

> states$State <- stri\_trans\_totitle(states$State)

> df <- data.frame(state.x77,

State = state.name,

Abbrev = state.abb,

Region = state.region,

Division = state.division

)

> df2 <- merge(states,df,by="State")

> df2 <- df2[order(df2$order),]

> mid\_range <- function(x) mean(range(x,na.rm=TRUE))

> centres <- ddply(df2, .(Abbrev),

colwise(mid\_range,.(lat,long,Population)))

> gg <- function(Cols) {

df2$Cols <- df2[,Cols]

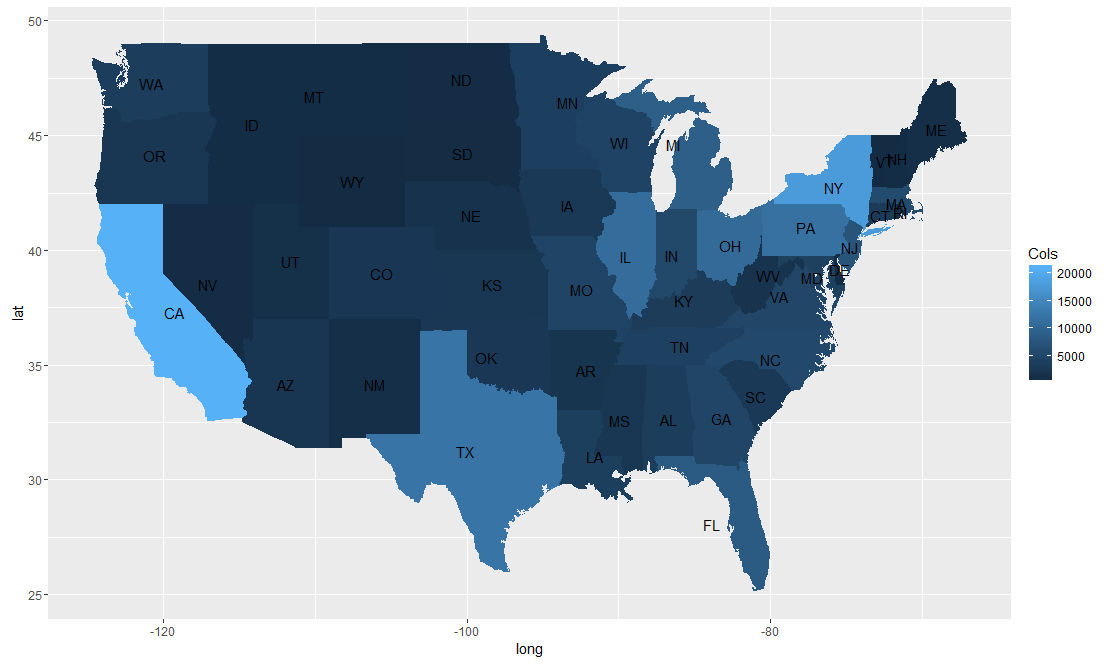
ggplot()+ + geom\_polygon(data = df2, aes(long,lat,fill=Cols,group=group)) +

geom\_text(aes(x=long,y=lat,label=Abbrev), data = centres, size=4)

}

>

> gg("Population")

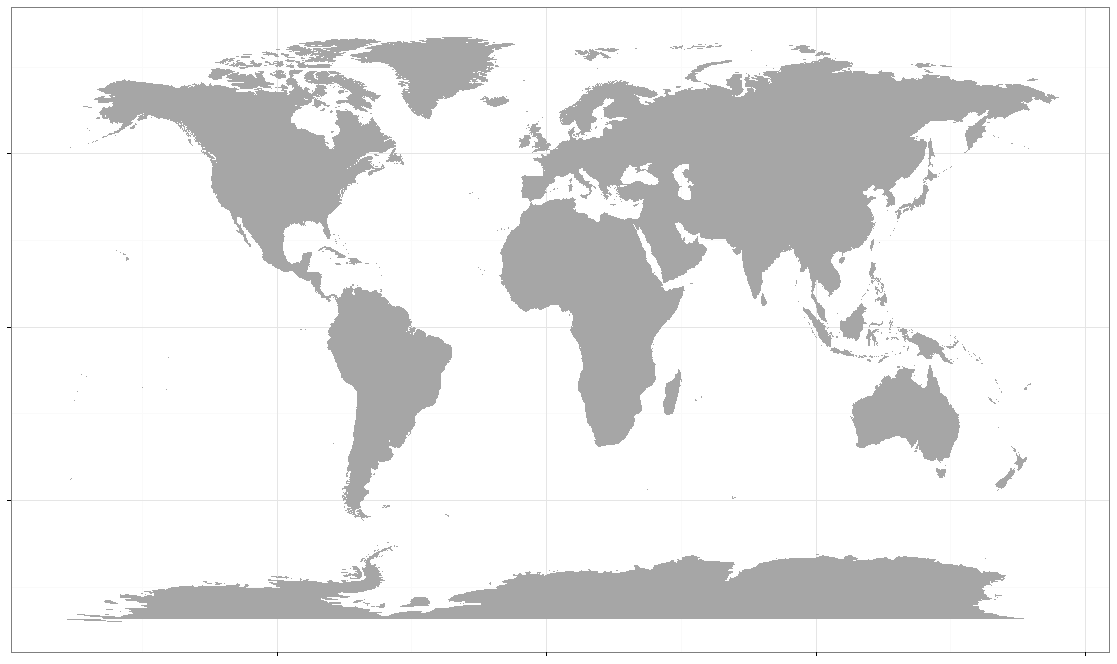


--- world map

> library(maps)

> map.dat <- map\_data("world")

> ggplot() + geom\_polygon(aes(long,lat, group=group), fill="grey65", data=map.dat) + theme\_bw() + theme(axis.text = element\_blank(), axis.title=element\_blank())



> ggplot(map.dat, aes(x=long, y=lat, group=group, fill=region))+ geom\_polygon() + theme(legend.position = "none")

