```r
rm(list = ls())
setwd("C:/Users/tuj53509/Dropbox/docs/Temple/Advanced Statistics for Urban Applications/Assignment1")
library(pastecs)
library(rgdal)
library(RColorBrewer)
library(GISTools)

#disable scientific notation
options(scipen=999)

#import data from a csv on kentucky broadband adoption by county
ky_bband_csv <- read.csv("KY.bband.csv", header=TRUE)
names(ky_bband_csv)

ky_bband_csv$county
#add leading zeros and toss it into a new vector to fix match with shapefile
ky_bband_csv$county2 <- formatC(ky_bband_csv$county, width = 3, format = "d", flag = "0")
#import spatial census data for kentucky from a shapefile
ky_tracts <- readOGR(dsn="ky_counties", layer="ky_counties")
plot(ky_tracts)
#add data from csv to the shape file's dataframe
ky_tracts@data <- data.frame(ky_tracts@data, ky_bband_csv[match(ky_tracts@data[,"COUNTYFP10"],
ky_bband_csv[,"county2"]),])

#attach to the data frame to simplify calling vectors
attach(data.frame(ky_tracts))

#get some basic descriptive statistics for many variables
stat.desc(ky_tracts@data[c("adopt05", "hhpct_0401", "medhhinc", "college", "perwhite", "unemp", "medage")])
#and specific to the variables in question
vals <- cbind(adopt05, college, hhpct_0401, popden2000)
stat.desc(vals)
head(adopt05)

#plot(adopt05)
pdf(file = "adopt05histogram.pdf")
hist(adopt05, col = "lightgreen", xlim = c(0,1), main = "Figure 1: Broadband Adoption by County", xlab = "broadband
adoption proportion")
dev.off()
pdf(file = "adopt05box.pdf")
boxplot(adopt05, hhpct_0401, col = "wheat3", names = c("adoption", "availability"), main = "Figure 2: Broadband
adoption vs. availability")
dev.off()

#make a choropleth map of adopt05 data
par(mar = c(0,0,0,0))
pdf(file="adopt05choropleth.pdf")
availability.shady <- auto.shading(adopt05, cols=brewer.pal(5,"Greens"))
choropleth(ky_tracts, adopt05, shading = availability.shady)
title("Figure 3: Kentucky Broadband Adoption Rate")
choro.legend(3751596,4833529,availability.shady, title = "proportion of broadband adoption")
map.scale(5021404, 3205804, miles2ft(200), "Miles", 5, 40)
north.arrow(5898350, 4350232, miles2ft(10))
```

```
dev.off()

#models
model1 <- lm(adopt05 ~ medhhinc)
model2 <- lm(adopt05 ~ medhhinc + college)
model3 <- lm(adopt05 ~ medhhinc + college + medage)
model4 <- lm(adopt05 ~ college + medage + hhpct_0401 + popden2000)
model5 <- lm(adopt05 ~ college + medage + hhpct_0401 + awareck05 + perwhite + noneed05)
model6 <- lm(adopt05 ~ college + hhpct_0401 + popden2000)
summary(model1)
summary(model2)
summary(model3)
summary(model4)
summary(model5)
summary(model6)
    #post regression diagnostics
pdf(file = "model6diags.pdf")
par(mar = c(3,4,4,2))
par(mfrow = c(2,2))
plot(model6)
dev.off()

#don't forget to detach the data frame
detach(data.frame(ky_tracts))
#reenable scientific notation
options(scipen=0)
```