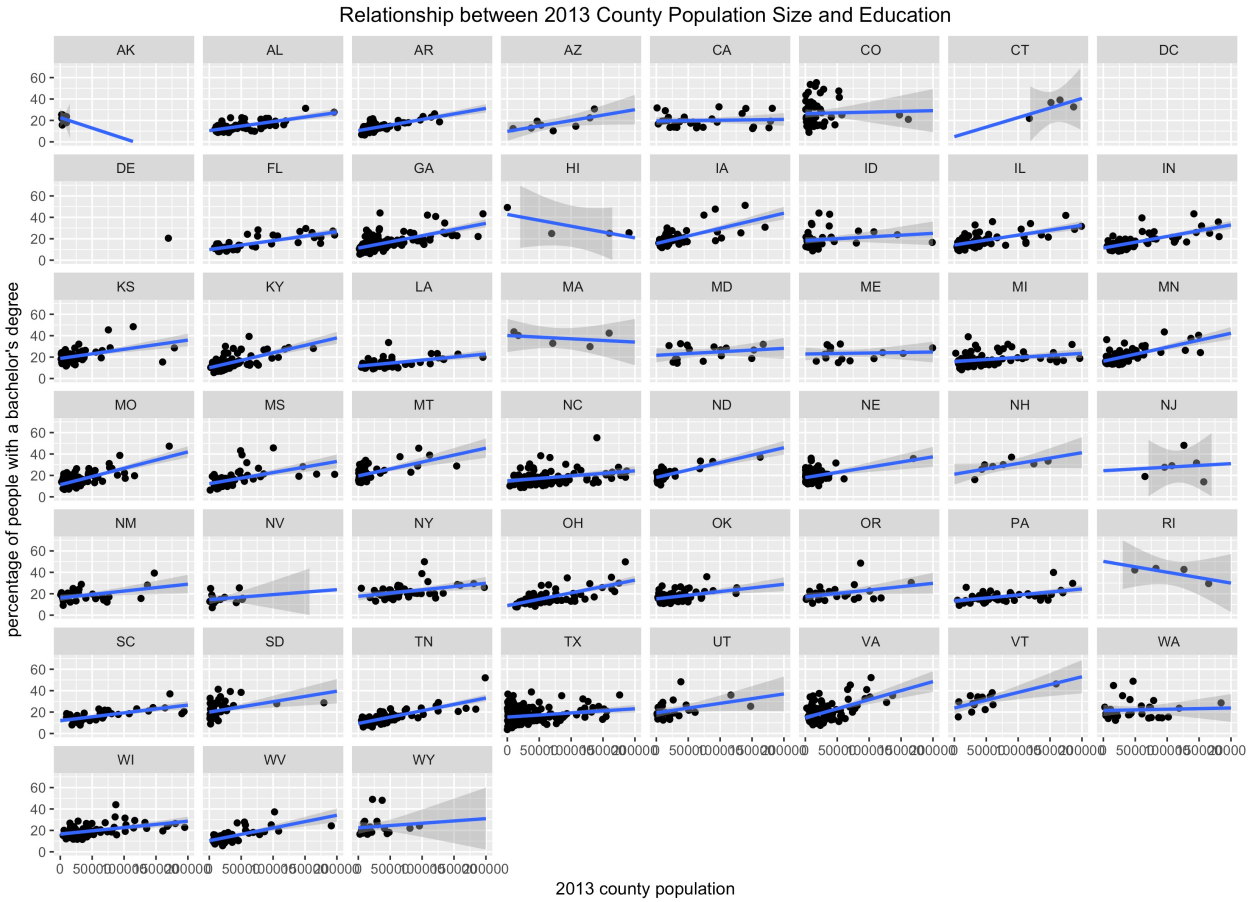
```
#Yang Yi
#10/06/2017
#HW7

#Install and use ggplot2
#install.packages("ggplot2")
library("ggplot2")

#Import data
census <- read.table('http://www.cse.lehigh.edu/~brian/census/DataSet.txt', sep=",",header =
TRUE)
reference <- read.fwf("http://www.cse.lehigh.edu/~brian/census/DataDict.txt", widths=c(10,
87, 4, 8, 11, 8, 10, 7),
col.names=c("Data_Item","Item_Description","Unit","Decimal","Us_Total","Minimum","Maxim
um","Source"))
county <- read.fwf('http://www.cse.lehigh.edu/~brian/census/FIPS_CountyName.txt',
widths=c(6, 50),col.names=c("flips", "counties"), stringsAsFactors=FALSE)

#Data cleaning, choose only those useful data
census <- subset(census, select = c("fips", "PST045213", "EDU685212"))
colnames(census) <- c("fips","population","education")
census$county <- county$counties

#Substitute the first column with only there states, and remove state rows
census$county <- gsub(".*,", "", census$county)
odd <- census$fips %% 2 != 0
census<-census[odd,]

#Factor county column
census$county <- factor(census$county)

#Plot population vs. education
plot <- qplot(data=census, x=population, y=education, facets=~county,xlim = c(0,200000), ylim
= c(0, 70)) + stat_smooth(data = census, method='lm', fullrange=TRUE)  + xlab("2013 county
population") + ylab("percentage of people with a bachelor's degree") + ggtitle("Relationship
between 2013 County Population Size and Education")+theme(plot.title = element_text(hjust =
0.5))

#Save the plot
ggsave("HW7.jpg")
```


Relationship between 2013 County Population Size and Education
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY
percentage of people with a bachelor's degree
2013 county population