```r
# Re-estimates the percent standard error of specific occupational employment statistics in
# metropolitan statistical areas (MSA's), using BLS data (http://www.bls.gov/oes/tables.htm),
# and then calculates a margin of error (upper and lower bounds) for those MSA's

# By Andrew Flowers <andrew.flowers@fivethirtyeight.com>
# See also http://fivethirtyeight.com/datalab/where-are-americas-librarians/

# install.packages(c("ggplot2", "stats"))
library(ggplot2)
library(stats)

# Loan and clean data
libMSA<-read.csv("librarians-by-msa.csv", header=T) # May 2013 data from BLS (http://www.bls.gov/oes/tables.htm)
names(libMSA)<-tolower(names(libMSA))
libMSA$tot_emp<-as.numeric(gsub("[$]|,", "", libMSA$tot_emp))
libMSA$emp_prse<-as.numeric(gsub("[$]|,", "", libMSA$emp_prse))

# Exploratory plots
plot(libMSA$tot_emp, libMSA$emp_prse, main="Librarian Employment vs. Standard Error", xlab="Librarian Employment", ylab="Standard Error")

# Linear model
l.model<-lm(libMSA$emp_prse~libMSA$tot_emp)
abline(l.model, col="red")
summary(l.model)   ### Linear model is bad

# Non-linear model
nl.model<-nls(formula=emp_prse~a*tot_emp^b, start=list(a=1, b=1), data=libMSA)
summary(nl.model)
a<-coef(nl.model)[1]; b<-coef(nl.model)[2]

plot(libMSA$tot_emp, libMSA$emp_prse, main="Librarian Employment vs. Standard Error", xlab="Librarian Employment", ylab="Standard Error")
curve(a*x^b, col='red', add=T)

# ggplot2 non-linear model
g<-ggplot(libMSA, aes(x=tot_emp, y=emp_prse))
g<-g+stat_smooth(method="nls", formula=y~a*x^b, se=FALSE, start=list(a=1, b=1))+geom_point()
g+ggtitle("Librarian Employment vs Standard Error")+ylab("Standard Error")+xlab("Librarian Employment")

# Create high and low estimates using new margin of error
libMSA$mor<-(a*(libMSA$tot_emp^b))*1.96
libMSA$high_emp<-libMSA$tot_emp*(1+(libMSA$mor/100))
```