

```
1 # Re-estimates the percent standard error of specific occupational employment statistics
  in
2 # metropolitan statistical areas (MSA's), using BLS data
  (http://www.bls.gov/oes/tables.htm),
3 # and then calculates a margin of error (upper and lower bounds) for those MSA's
4
5 # By Andrew Flowers <andrew.flowers@fivethirtyeight.com>
6 # See also http://fivethirtyeight.com/datalab/where-are-americas-librarians/
7
8 # install.packages(c("ggplot2", "stats"))
9 library(ggplot2)
10 library(stats)
11
12 # Load and clean data
13 libMSA<-read.csv("librarians-by-msa.csv", header=T) # May 2013 data from BLS
  (http://www.bls.gov/oes/tables.htm)
14 names(libMSA)<-tolower(names(libMSA))
15 libMSA$tot_emp<-as.numeric(gsub("[^$]|", "", libMSA$tot_emp))
16 libMSA$emp_prse<-as.numeric(gsub("[^$]|", "", libMSA$emp_prse))
17
18 # Exploratory plots
19 plot(libMSA$tot_emp, libMSA$emp_prse, main="Librarian Employment vs. Standard Error",
  xlab="Librarian Employment", ylab="Standard Error")
20
21 # Linear model
22 l.model<-lm(libMSA$emp_prse~libMSA$tot_emp)
23 abline(l.model, col="red")
24 summary(l.model) ### Linear model is bad
25
26 # Non-Linear model
27 nl.model<-nls(formula=emp_prse~a*tot_emp^b, start=list(a=1, b=1), data=libMSA)
28 summary(nl.model)
29 a<-coef(nl.model)[1]; b<-coef(nl.model)[2]
30
31 plot(libMSA$tot_emp, libMSA$emp_prse, main="Librarian Employment vs. Standard Error",
  xlab="Librarian Employment", ylab="Standard Error")
32 curve(a*x^b, col='red', add=T)
33
34 # ggplot2 non-linear model
35 g<-ggplot(libMSA, aes(x=tot_emp, y=emp_prse))
36 g<-g+stat_smooth(method="nls", formula=y~a*x^b, se=FALSE, start=list(a=1,
  b=1))+geom_point()
37 g+ggtitle("Librarian Employment vs Standard Error")+ylab("Standard
  Error")+xlab("Librarian Employment")
38
39 # Create high and low estimates using new margin of error
40 libMSA$mor<-(a*(libMSA$tot_emp^b))*1.96
41 libMSA$high_emp<-libMSA$tot_emp*(1+(libMSA$mor/100))
```