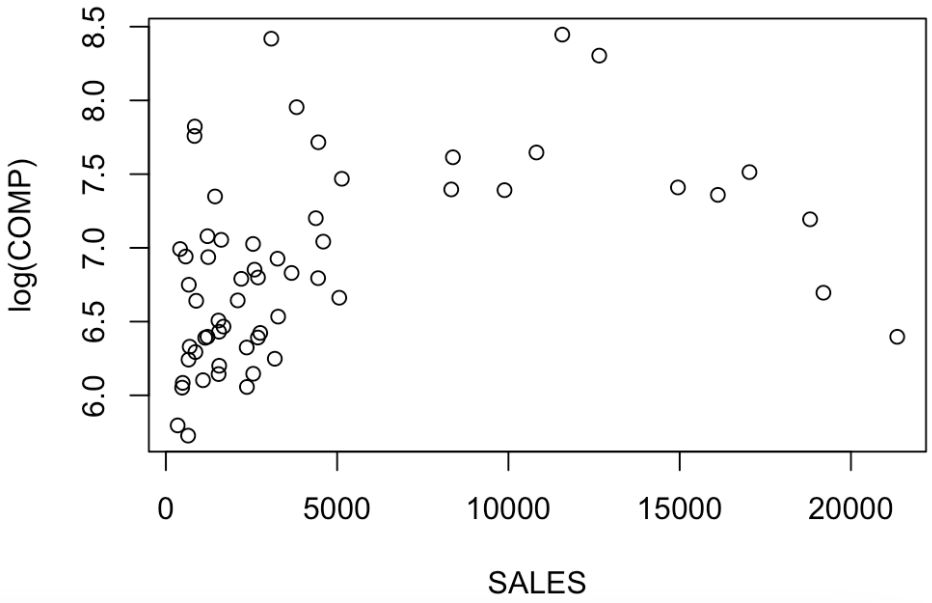
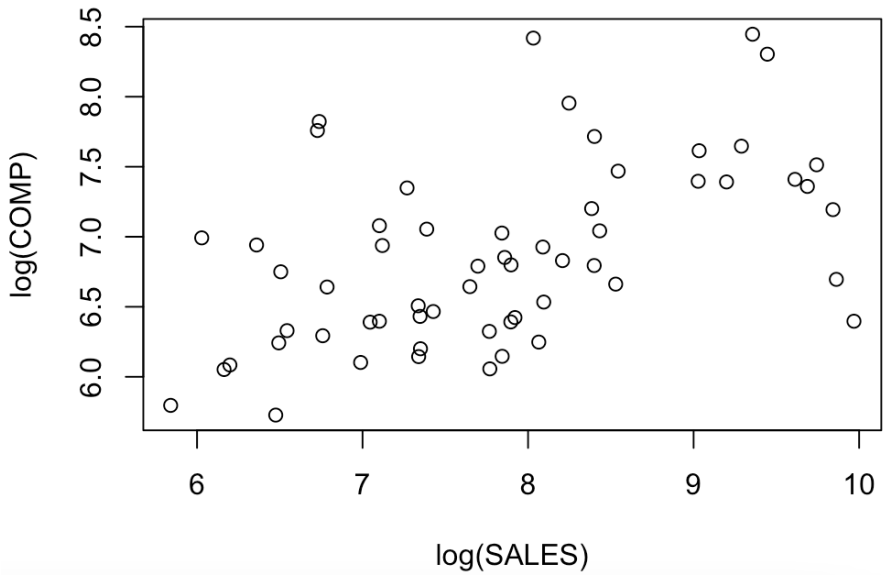
3) a) By observing the scatterplot of SALES vs log(COMP), a linear model seems inappropriate for these two variates since there seems like a quadratic function pattern between these two variates.

```
# 3)
# a)
plot(SALES,log(COMP))
```



b) By observing the scatterplot of log(SALES) vs log(COMP), a linear model seems appropriate for these two variates since there is no observable pattern. Thus, the scatterplot in part b is more like a linear model compared with the scatterplot in part a.

```
# b)
plot(log(SALES),log(COMP))
```



c) These transformations make the model better. To be specific, based on the R-code output, we find that the R-squared value is 0.5317, which is bigger than 0.4883 and 0.485. Thus, this represents that approximately 53.17% of the variation of the log transformation of CEO compensation can be explained by the model variables, including CEO's age, CEO's education level, background type, number of years employed by the firm, number of years as the firm CEO, the log transformation of sales revenues, the log transformation of market value of the CEO's stock, the log transformation of percentage of firm's market value owned by the CEO, the log transformation of profits of the firm before taxes.

Moreover, we find that the p-values of log(VAL) and log(PCNTOWN) are smaller than 0.05. After accounting for the other variables in the model, the log transformation of market value of the CEO's stock and the log transformation of percentage of firm's market value owned by the CEO are significantly related to the log transformation of CEO compensation. Other variables, including CEO's age, CEO's education level, background type, number of years employed by the firm, number of years as the firm CEO, the log transformation of sales revenues, the log transformation of profits of the firm before taxes, are not significantly related to the log transformation of CEO compensation after accounting for other variables in the model since their p-values are larger than 0.05 based on the R code output.

```
# c)
refmagain <- lm(log(COMP) ~ AGE+EDUCATN+bg+TENURE+EXPER+log(SALES)+log(VAL)+
                            log(PCNTOWN)+log(PROF))
summary(refmagain)
```

```
Call:
lm(formula = log(COMP) ~ AGE + EDUCATN + bg + TENURE + EXPER +
    log(SALES) + log(VAL) + log(PCNTOWN) + log(PROF))

Residuals:
     Min       1Q   Median       3Q      Max
-0.95958 -0.32822  0.01758  0.18959  1.37367

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.362003   0.995175   5.388 2.37e-06 ***
AGE           0.007900   0.013590   0.581  0.56386
EDUCATN      -0.118305   0.167322  -0.707  0.48310
bg2           0.035371   0.222855   0.159  0.87458
bg3           0.285768   0.225508   1.267  0.21145
bg4          -0.090375   0.245194  -0.369  0.71413
bg5           0.078269   0.214047   0.366  0.71629
TENURE        0.004490   0.007748   0.579  0.56508
EXPER        -0.002009   0.011396  -0.176  0.86084
log(SALES)    0.097538   0.093626   1.042  0.30296
log(VAL)      0.421421   0.120425   3.499  0.00105 **
log(PCNTOWN) -0.366457   0.129754  -2.824  0.00698 **
log(PROF)    -0.191712   0.113117  -1.695  0.09687 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4969 on 46 degrees of freedom
Multiple R-squared:  0.5317,    Adjusted R-squared:  0.4095
F-statistic: 4.352 on 12 and 46 DF,  p-value: 0.0001339
```