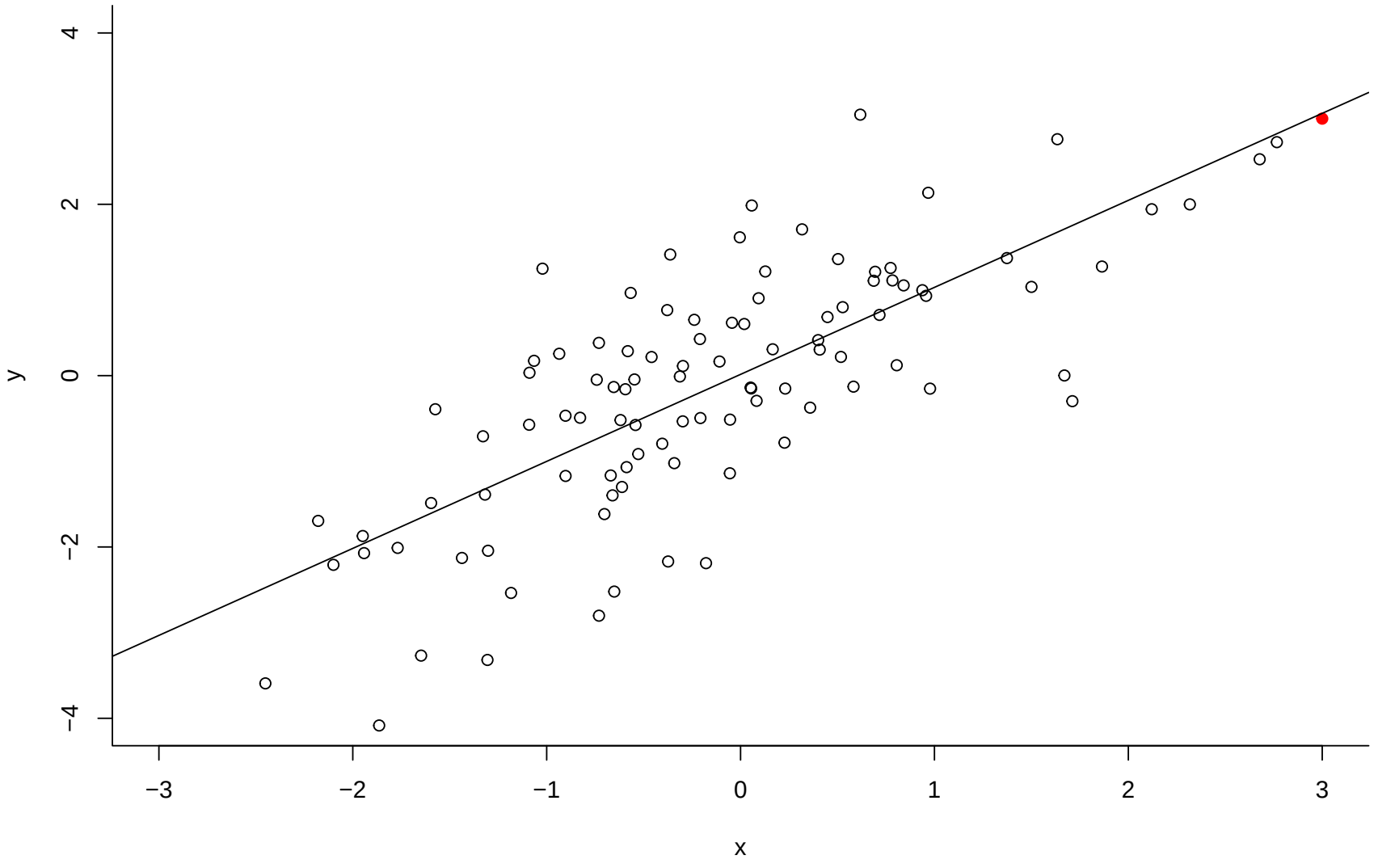
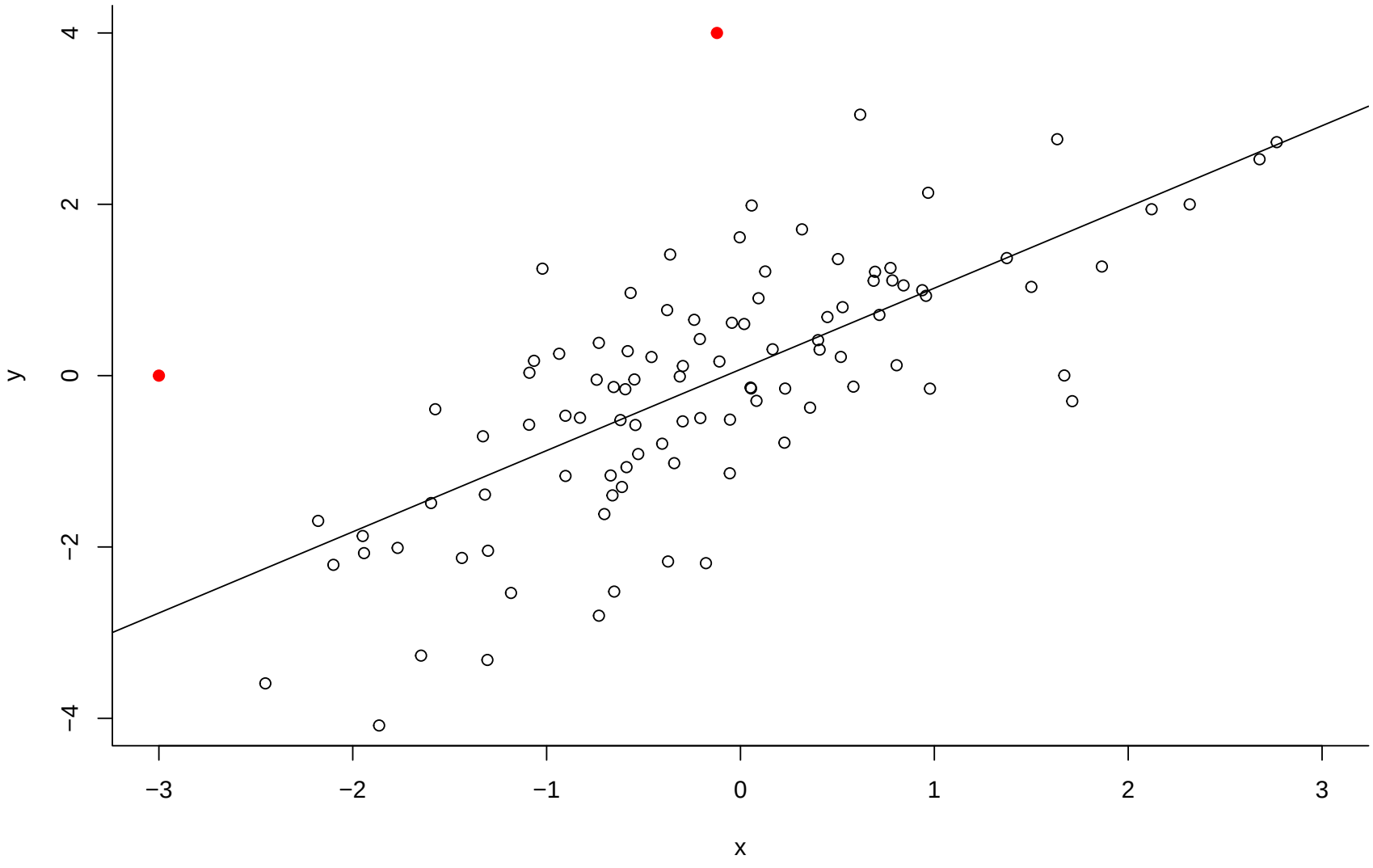
Outliers and Influence in Linear Regression

Randy Johnson

Influence



Outliers



Outliers

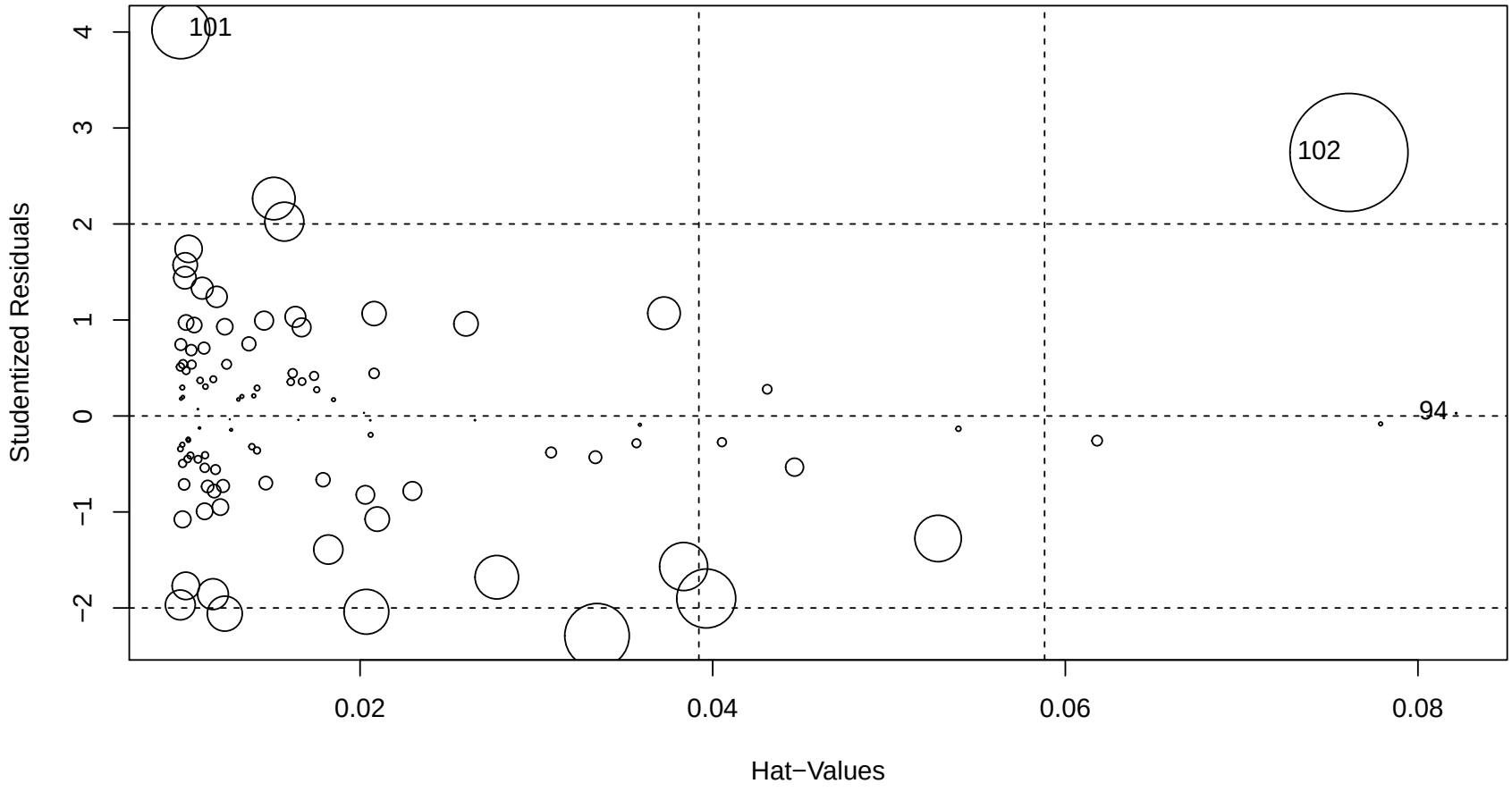
```
lm0 <- lm(y ~ x, data = dat)

require(car)
outlierTest(lm0)
```

```
##     rstudent unadjusted p-value Bonferonni p
## 101 4.021705         0.00011296     0.011522
```

```
outlierTest(lm0, cutoff=0.8)
```

```
##     rstudent unadjusted p-value Bonferonni p
## 101 4.021705         0.00011296     0.011522
## 102 2.745099         0.00718480     0.732850
```

Leverage




```
##          StudRes         Hat        CookD
## 94   0.02878472 0.082168292 3.746248e-05
## 101  4.02170469 0.009832767 6.972728e-02
## 102  2.74509866 0.076088424 2.912591e-01
```

Outliers and Leverage

- Residuals: The difference between the observed and predicted value of the outcome variable.
- Studentized residuals: Scaled residuals, such that the variance $= 1$.
- Hat-values: A measure of the amount of influence each data point has on the outcome predictions.
- Cook's distance: A measure of the effect of each data point on the regression coefficients.