Polynomial Regression -- Splus Commands and Output

```
> plotresids <- function(linmod)
{
    plot(linmod$fitted, linmod$residuals)
    title("Residuals vs. Fitted")
    abline(0, 0)
    qqnorm(linmod$residuals)
    title("QQ Normal Plot")
}

> attach(allele)

> #linear model
lm.linear <- lm(Frequency ~ Distance)
> par(mfrow = c(4, 3))
> plot(Distance, Frequency)
> abline(lm.linear$coefficients[1], lm.linear$coefficients[2])
> title("Linear Model")
> plotresids(lm.linear)

> #quadratic model
lm.quad <- lm(Frequency ~ Distance + Distance^2)
> plot(Distance, Frequency)
> title("Quadratic Model")
> dist <- seq(from = 1, to = 69, length = 100)
> dist.mat <- as.matrix(cbind(1, dist, dist^2))
> freq <- dist.mat %*% lm.quad$coefficients
> lines(dist, freq)
> plotresids(lm.quad)

> #cubic model
lm.cubic <- lm(Frequency ~ Distance + Distance^2 + Distance^3)
> plot(Distance, Frequency)
> title("Cubic Model")
> dist <- seq(from = 1, to = 69, length = 100)
> dist.mat <- as.matrix(cbind(1, dist, dist^2, dist^3))
> freq <- dist.mat %*% lm.cubic$coefficients
> lines(dist, freq)
> plotresids(lm.cubic)
> print(summary(lm.linear))

Call: lm(formula = Frequency ~ Distance)

Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 0.0458 0.0364     1.2604  0.2268
   Distance 0.0073 0.0009     8.4204  0.0000

Residual standard error: 0.06955 on 15 degrees of freedom
Multiple R-Squared: 0.8254
F-statistic: 70.9 on 1 and 15 df, the p-value is 4.557e-007

> print(summary(lm.quad))
Call: lm(formula = Frequency ~ Distance + Distance^2)

Coefficients:
               Value Std. Error t value Pr(>|t|)
  (Intercept) 0.1173 0.0480     2.4409  0.0285
     Distance 0.0008 0.0032     0.2618  0.7973
I(Distance^2) 0.0001 0.0000     2.0470  0.0599

Residual standard error: 0.06316 on 14 degrees of freedom
Multiple R-Squared: 0.8656
F-statistic: 45.09 on 2 and df, the p-value is 7.919e-007

> print(summary(lm.cubic))

Call: lm(formula = Frequency ~ Distance + Distance^2 + Distance^3)

Coefficients:
                Value Std. Error t value Pr(>|t|)
  (Intercept)  0.2010  0.0547     3.6761  0.0028
     Distance -0.0137  0.0068    -2.0277  0.0636
I(Distance^2)  0.0006  0.0002     2.7554  0.0164
I(Distance^3)  0.0000  0.0000    -2.3662  0.0342

Residual standard error: 0.0548 on 13 degrees of freedom
Multiple R-Squared: 0.9061
F-statistic: 41.8 on 3 and 13 df, the p-value is 6.134e-007

> ## adding higher order polynomials comes at a price
## let's look at collinearity in this polynomial model:
print(summary(lm(Distance ~ Distance^2))$r.squared)
[1] 0.9412423
> print(summary(lm(Distance ~ Distance^2 + Distance^3))$r.squared)
[1] 0.9898842

> ## bias/variance tradeoff?
aic.bic.fun <- function(n, p, s2)
{
    AIC <- n * log(s2) + 2 * p
    BIC <- n * log(s2) + p * log(n)
    print(paste("AIC=", round(AIC, 2)))
    print(paste("BIC=", round(BIC, 2)))
}
> n <- 17
> aic.bic.fun(n, 2, (summary(lm.linear)$sigma^2))
[1] "AIC= -86.63"
[1] "BIC= -84.97"

> aic.bic.fun(n, 3, (summary(lm.quad)$sigma^2))
[1] "AIC= -87.91"
[1] "BIC= -85.41"

> aic.bic.fun(n, 4, (summary(lm.cubic)$sigma^2))
[1] "AIC= -90.74"
[1] "BIC= -87.41"

> detach(what = 2)
```

Polynomial Regression -- Splus Commands and Output