1.

$$MY = Y - X(X'X)^{-1}X'Y$$

$$= X\beta + e - X(X'X)^{-1}X'(X\beta + e)$$

$$= X\beta + e - X\beta - X(X'X)^{-1}X'e$$

$$= Me = \hat{e}$$

$$C_p = \hat{e}'\hat{e} + 2ks^2 = \frac{n+k}{n-k}\cdot\hat{e}'\hat{e} \Rightarrow E[C_p] = \frac{n+k}{n-k}\cdot E[\hat{e}'\hat{e}]$$

$$E[\hat{e}'\hat{e}] = E[E[\hat{e}'\hat{e}\,|\,X]] = E[n\cdot\sigma^2(1-\tfrac{k}{n})] = (n-k)\sigma^2$$

$$\Rightarrow E[C_p] = (n+k)\sigma^2$$

$$R = E[\|\hat{m} - m\|^2]$$

$$= E[(X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta)]$$

$$= E[(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)] = \text{trace}(X'X\cdot\sigma^2(X'X)^{-1}) = \sigma^2\cdot\text{trace}(I_k) = k\sigma^2$$

(covariance matrix of $\hat{\beta}-\beta$; $\hat{\beta}-\beta \sim N(0, \sigma^2(X'X)^{-1})$)

$$\Rightarrow E[C_p] = n\sigma^2 + k\sigma^2 = n\sigma^2 + R \quad \#$$

2.1

$$\hat{e}_1 = Y - \hat{m}_1 = Y - X_1\hat{\beta}_1 = X_1\beta_1 + X_2\beta_2 + e - X_1(X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + e)$$

$$= M_1X_2\beta_2 + M_1e \quad \text{with } M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$$

$$\hat{e}_1'\hat{e}_1 = \beta_2'X_2'M_1^2X_2\beta_2 + \beta_2'X_2'M_1^2e + e'M_1^2X_2\beta_2 + e'M_1^2e \quad \text{and } M_1^2 = M_1$$

$$= \beta_2'X_2'M_1X_2\beta_2 + \beta_2'X_2'M_1e + e'M_1X_2\beta_2 + e'M_1e$$

$$E[\hat{e}_1'\hat{e}_1\,|\,X] = \beta_2'X_2'M_1X_2\beta_2 + 0 + 0 + \sigma^2\text{trace}(M_1) = \beta_2'X_2'M_1X_2\beta_2 + (n-k_1)\sigma^2$$

$$E[C_{1,p}] = E[\hat{e}_1'\hat{e}_1 + 2k_1s^2] = E[\hat{e}_1'\hat{e}_1] + 2k_1\cdot E[s^2] = E[E[\hat{e}_1'\hat{e}_1|X]] + 2k_1\sigma^2$$

$$\Rightarrow E[C_{1,p}] = E[\beta_2'X_2'M_1X_2\beta_2 + (n-k_1)\sigma^2] + 2k_1\sigma^2 = E[\beta_2'X_2'M_1X_2\beta_2] + (n+k_1)\sigma^2 = R + n\sigma^2 \quad \#$$

P3

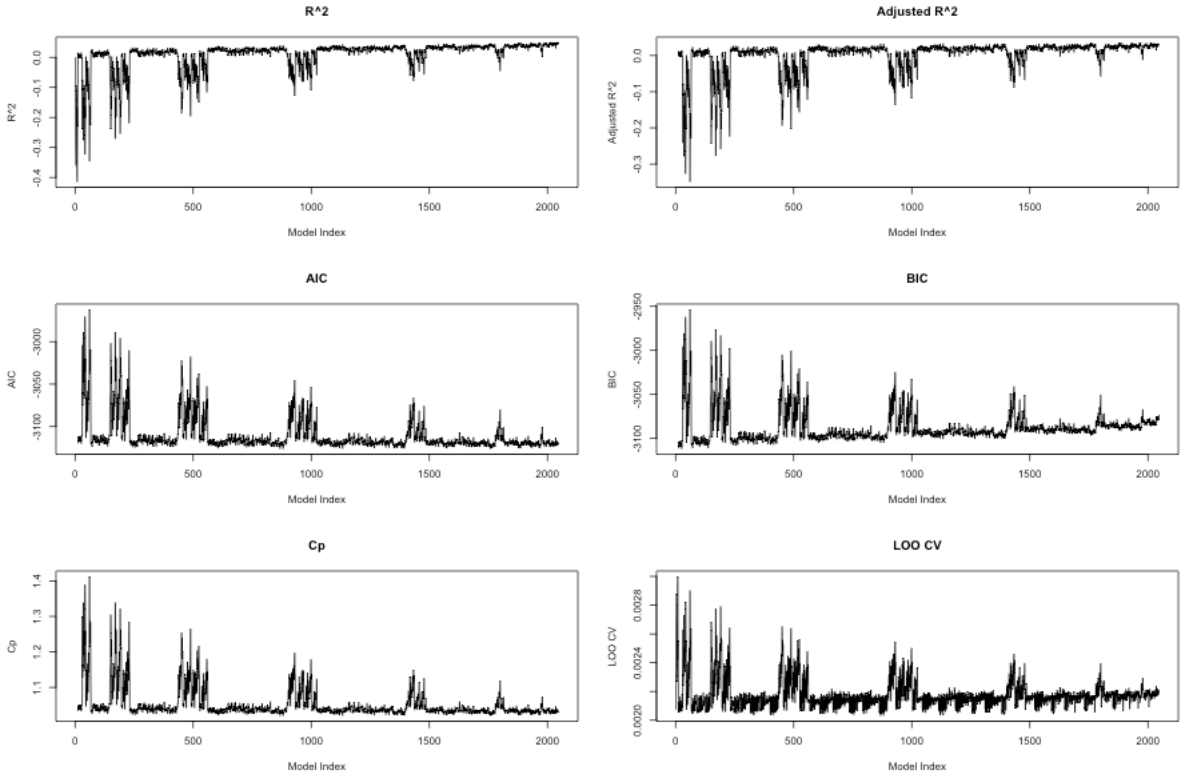
Compute each model's criterian

```r
# Define the function to collect criteria
collect_criterian <- function(y, x) {
  n <- length(y)
  k <- ncol(x)
  x_matrix <- as.matrix(x)
  beta_hat <- solve(t(x_matrix) %*% x_matrix) %*% t(x_matrix) %*% y
  y_hat <- x_matrix %*% beta_hat
  residuals <- y - y_hat
  rss <- sum(residuals^2)
  tss <- sum((y - mean(y))^2)
  sigma_hat <- rss / (n - k)
  r2 <- 1 - rss / tss
  adjusted_r2 <- 1 - ((1 - r2) * (n - 1) / (n - k))
  aic <- n * log(rss / n) + 2 * k
  bic <- n * log(rss / n) + k * log(n)
  cp <- rss + 2 * k * sigma_hat
  h_ii <- rowSums((x_matrix %*% solve(t(x_matrix) %*% x_matrix)) * x_matrix)
  loo_cv <- mean((residuals / (1 - h_ii))^2)

  model_results <- list(r2 = r2, adjusted_r2 = adjusted_r2, aic = aic,
                        bic = bic, cp = cp, loo_cv = loo_cv)
  return(model_results)
}
```

Results

Value of the criteria vs model index



Model index is ordered by coverting the parameter selection type into binary number. For instance, the model selecting only "ones" has index 1 and the model selecting "ones", "dfy", "dfy_square" has index 1+2+64=67.

Best model selected from each criteria

```
Best Model Based on R2:
Predictors:
 [1] "ones"           "x_dfy"          "x_infl"         "x_svar"         "x_tms"          "x_tbl"          "x_dfy_squared"  "x_infl_squared"
 [9] "x_svar_squared" "x_tms_squared"  "x_tbl_squared"
r2: 0.0493

Best Model Based on ADJUSTED_R2:
Predictors:
[1] "ones"           "x_dfy"          "x_svar"         "x_tms"          "x_infl_squared" "x_svar_squared" "x_tms_squared"  "x_tbl_squared"
adjusted_r2: 0.0358

Best Model Based on AIC:
Predictors:
[1] "x_dfy"          "x_tms"          "x_tbl"          "x_infl_squared" "x_tms_squared"
aic: -3126.1674

Best Model Based on BIC:
Predictors:
[1] "x_dfy"         "x_dfy_squared"
bic: -3111.6190

Best Model Based on CP:
Predictors:
[1] "x_dfy"          "x_tms"          "x_tbl"          "x_infl_squared" "x_tms_squared"
cp: 1.0201

Best Model Based on LOO_CV:
Predictors:
[1] "x_dfy"          "x_tms"          "x_tbl"          "x_infl_squared" "x_tms_squared"
loo_cv: 0.0020
```

Noted that AIC, Cp, LOOCV choose the same model as the best model.

Number of predictors that the best model selected

r2: 11 (predictors)
sdjusted_r2: 8
AIC: 5
BIC: 2
Cp: 5
Loo_cv: 5