Chap 5 homework

Exercise 3

(a)

1. divide the data set into k chunks, 1 for testing data and k-1 for training data
2. apply the model needed to validate, use k-1 chunks for training and 1 chunks for test, and take turns to get k test errors, then average it for the model's test error.
3. use next model (next k or other poly degree), repeat 2 to get another test error.

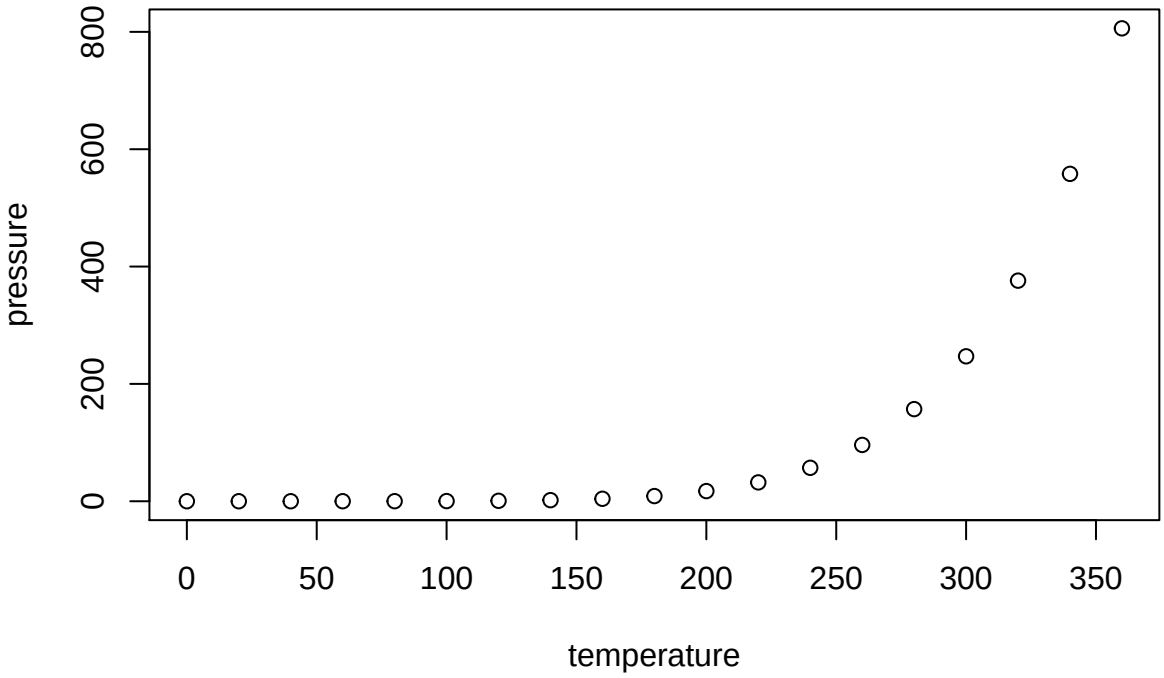
(b)

i. The pro of k-fold cv relative to the validation approach is that its test error is much more deterministic whereas the estimate of validation approach is highly variable. In addition, the validation's test error may overestimate the test error. In the other hand, the con of k-fold cv relative to the validation approach is that k-fold cv is much more computationally intensive.

ii.

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```


pressure
0
200
400
600
800
0
50
100
150
200
250
300
350
temperature