Result for lab3ex1.R for Exercise 1:

```
> dim(abalone.test)
[1] 1176   10
> dim(abalone.train)
[1] 3000   10
```

K=50

```
=== kNN Model A (length, diameter, height; k = 50) ===
> print(tab.A)
          actual
predicted young adult old
    young   278    76  13
    adult   112   392 152
    old      19    58  76
> cat("Accuracy A: ", round(acc.A, 4), "\n", sep = "")
Accuracy A: 0.6344
```

```
=== kNN Model B (whole_weight, shell_weight; k = 50) ===
> print(tab.B)
          actual
predicted young adult old
    young   296    86  13
    adult   108   383 142
    old       5    57  86
> cat("Accuracy B: ", round(acc.B, 4), "\n", sep = "")
Accuracy B: 0.6505
```

Find the optimal value *k* for the better performing model by training over k =1:100

```
=== k tuning for best model: B (weight features) ===
> cat("Best k: ", best.k, "\n", sep = "")
Best k: 86
> cat("Best tuned accuracy: ", round(best.acc, 4), "\n", sep = "")
Best tuned accuracy: 0.6641
```

Result for lab3findbestsubset.R to find the best subset for Exercise 2

```
> dim(abalone.test)
[1] 1176   10
> dim(abalone.train)
[1] 3000   10
```

```
=== kNN Model A (length, diameter, height; k = 50) ===
> print(tab.A)
          actual
predicted young adult old
    young   278    76  13
    adult   112   392 152
    old      19    58  76
> cat("Accuracy A: ", round(acc.A, 4), "\n", sep = "")
Accuracy A: 0.6344
```

```
=== kNN Model B (whole_weight, shucked_weight, viscera_weight, shell_weight; k = 50) ===
> print(tab.B)
          actual
predicted young adult old
    young   301    89  10
    adult   102   393 126
    old       6    44 105
> cat("Accuracy B: ", round(acc.B, 4), "\n", sep = "")
Accuracy B: 0.6794
> acc_from_tab <- function(tab) sum(diag(tab)) / sum(tab)
```

Find the optimal value *k* for Subset A and Subset B by training over k =1:100

```
Subset A (length, diameter, height)
> cat("  best acc :", round(resA$best_acc,4), " at k=", resA$best_k_acc,  "\n", sep="")
  best acc :0.6395 at k=52
> cat("\nSubset B (", paste(feats.B, collapse=", "), ")\n", sep="")

Subset B (whole_weight, shucked_weight, viscera_weight, shell_weight)
> cat("  best acc :", round(resB$best_acc,4), " at k=", resB$best_k_acc,  "\n", sep="")
  best acc :0.693 at k=15
> cat("\n>>> Best subset:", best_subset,
+     "with k =", best_k, "(by accuracy)\n")

>>> Best subset: B with k = 15 (by accuracy)
```

Result for lab3ex2.R for Exercise 2:

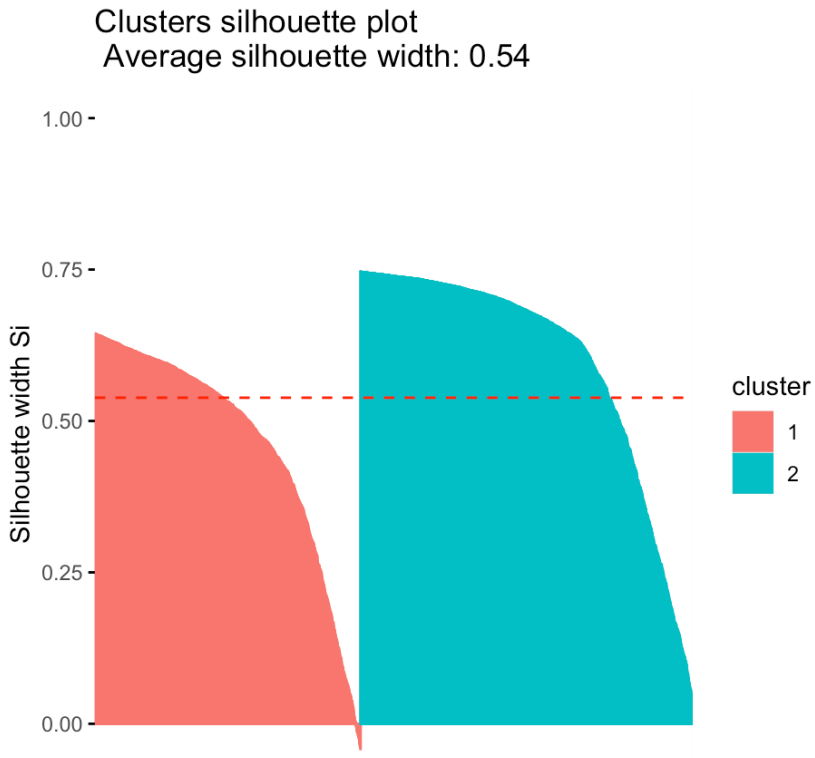
We choose subset (whole_weight,shucked_weight, viscera_weight, shell_weight) from best performing feature subset from the previous lab3findbestsubset.R.

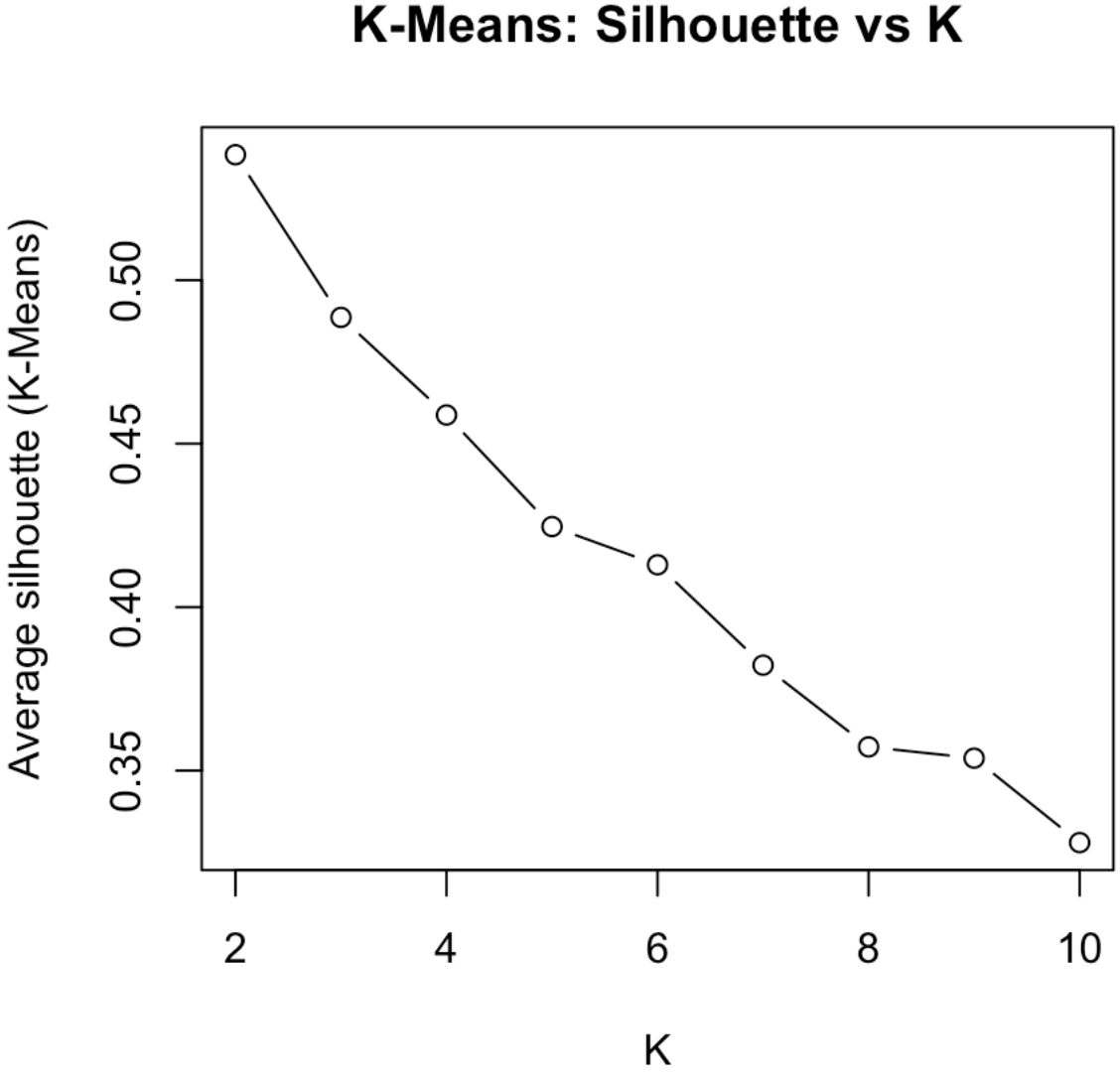
And for K-Means and PAM, we compare the result of the model with k=2:10 to select the best k.

For K-means:

```
> cat("Best K for K-Means by avg silhouette:", best.k.km, "\n")
Best K for K-Means by avg silhouette: 2
```

```
> fviz_silhouette(sil.km)
  cluster size ave.sil.width
1       1 1857          0.46
2       2 2319          0.60
```

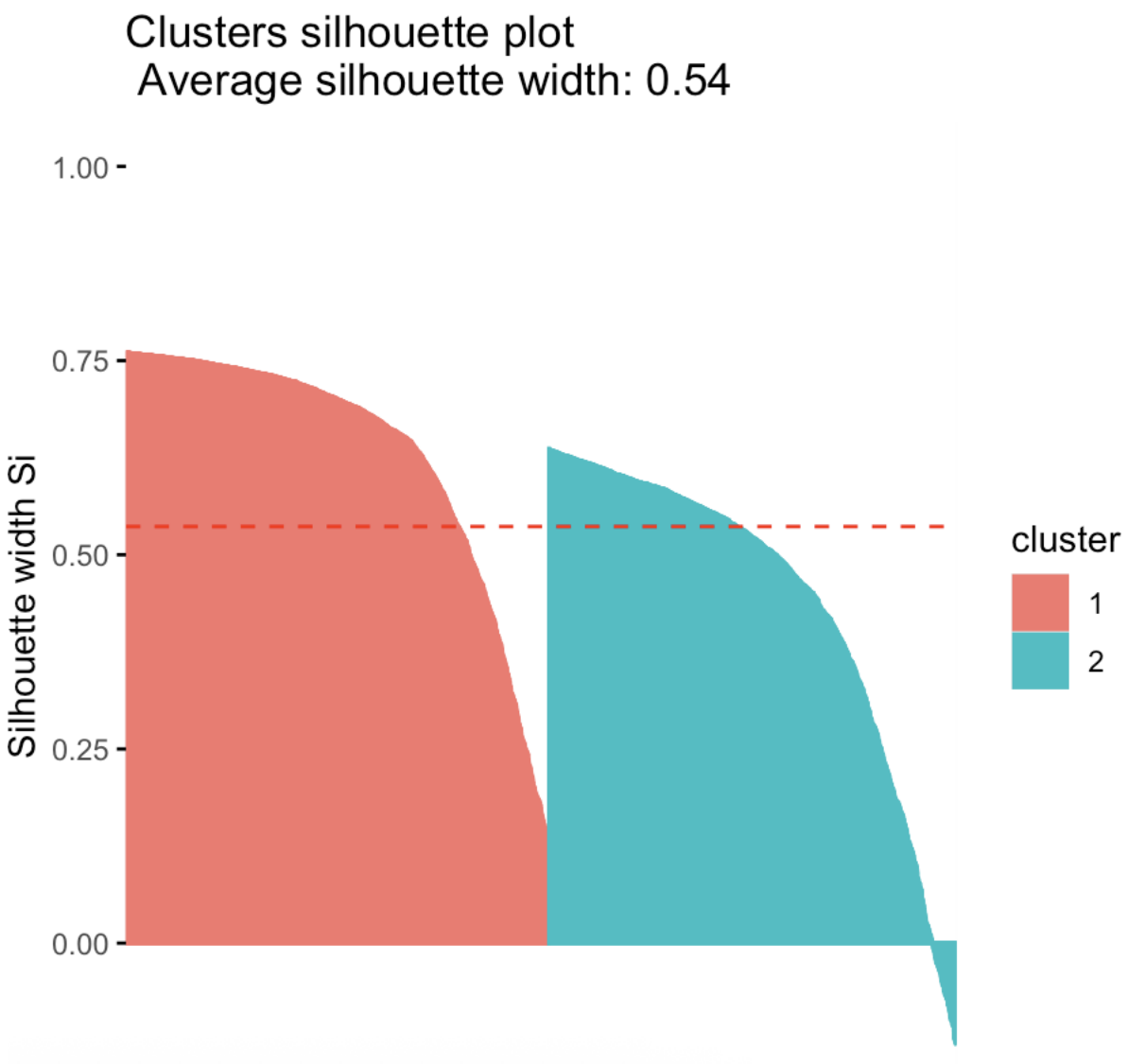


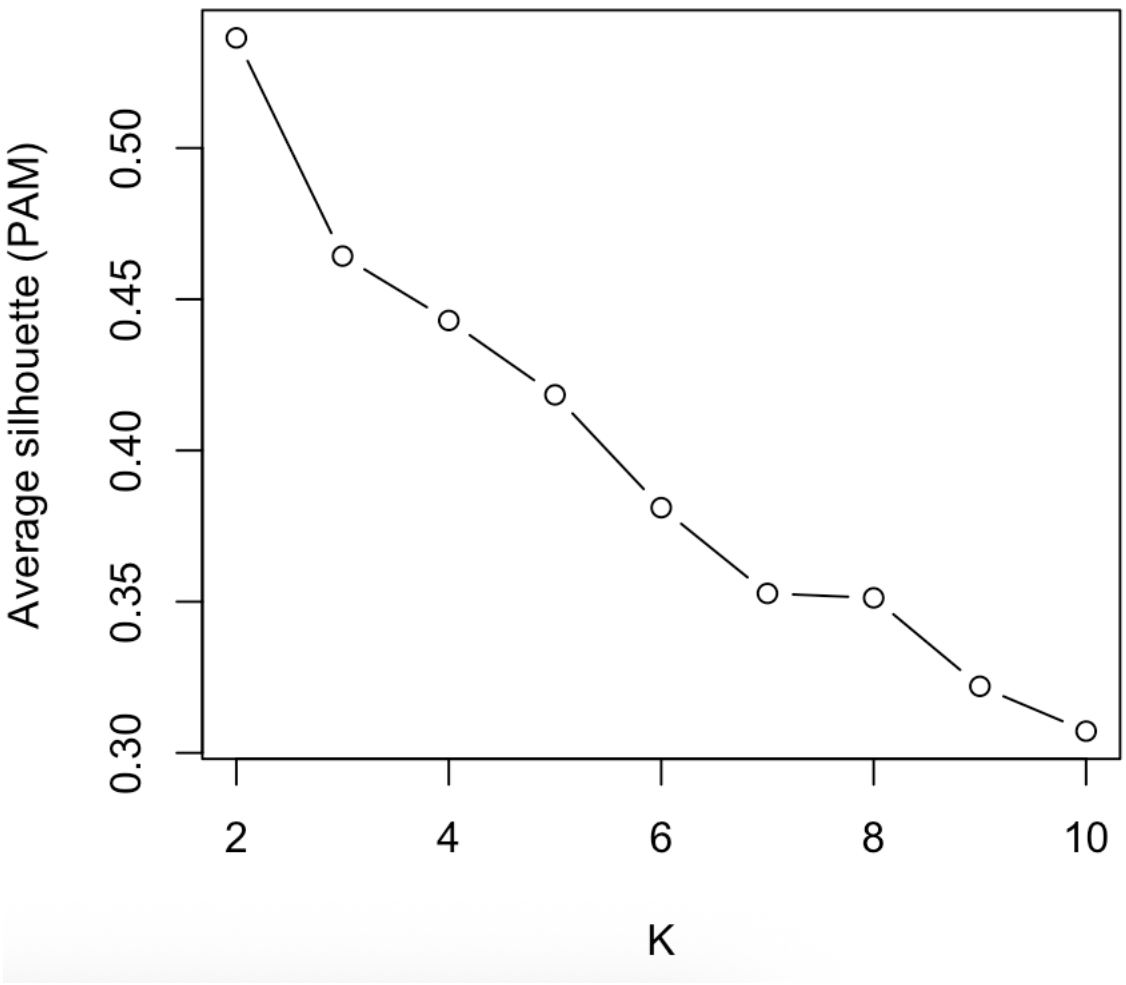

K-Means: Silhouette vs K
Average silhouette (K-Means)
0.35
0.40
0.45
0.50
2
4
6
8
10
K

For PAM

```
> cat("Best K for PAM by avg silhouette:", best.k.pam, "\n")
Best K for PAM by avg silhouette: 2
```

```
> fviz_silhouette(sil.pam)
  cluster size ave.sil.width
1       1 2127          0.63
2       2 2049          0.44
```




PAM: Silhouette vs K
Average silhouette (PAM)
0.30
0.35
0.40
0.45
0.50
2
4
6
8
10
K