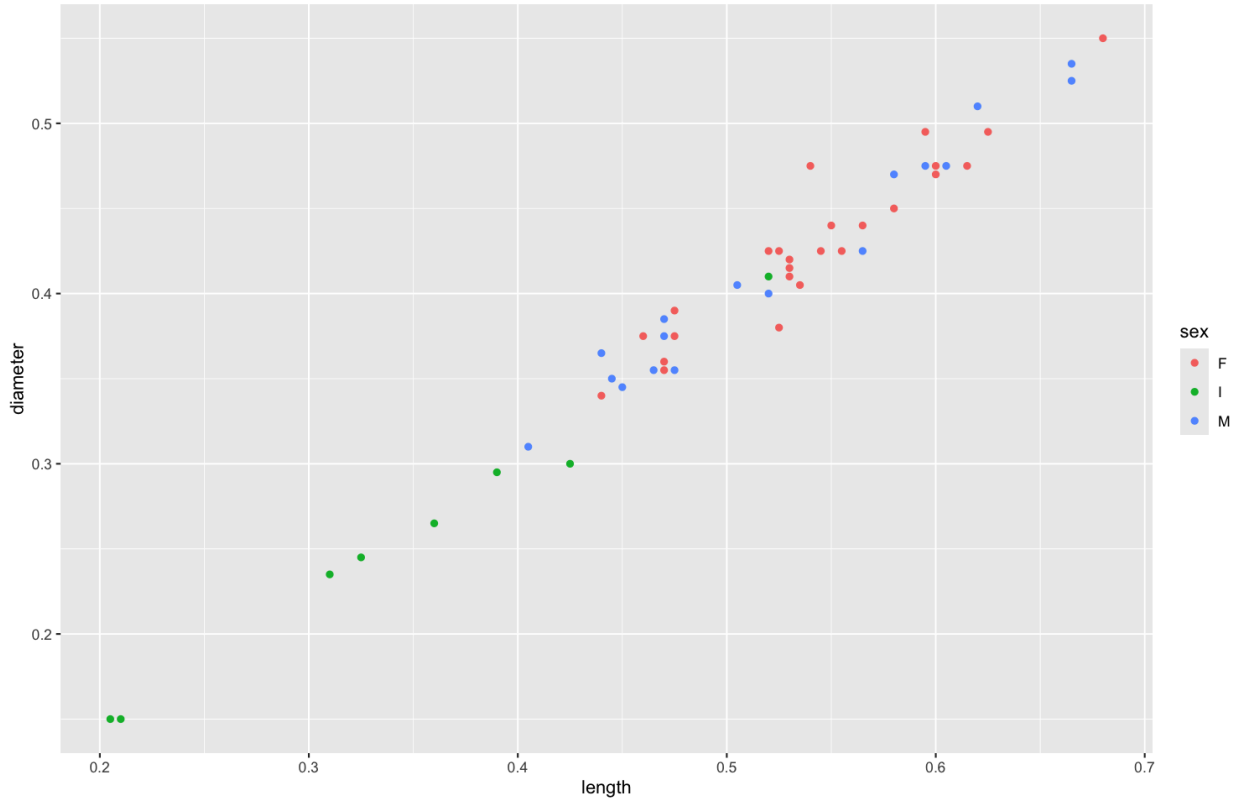
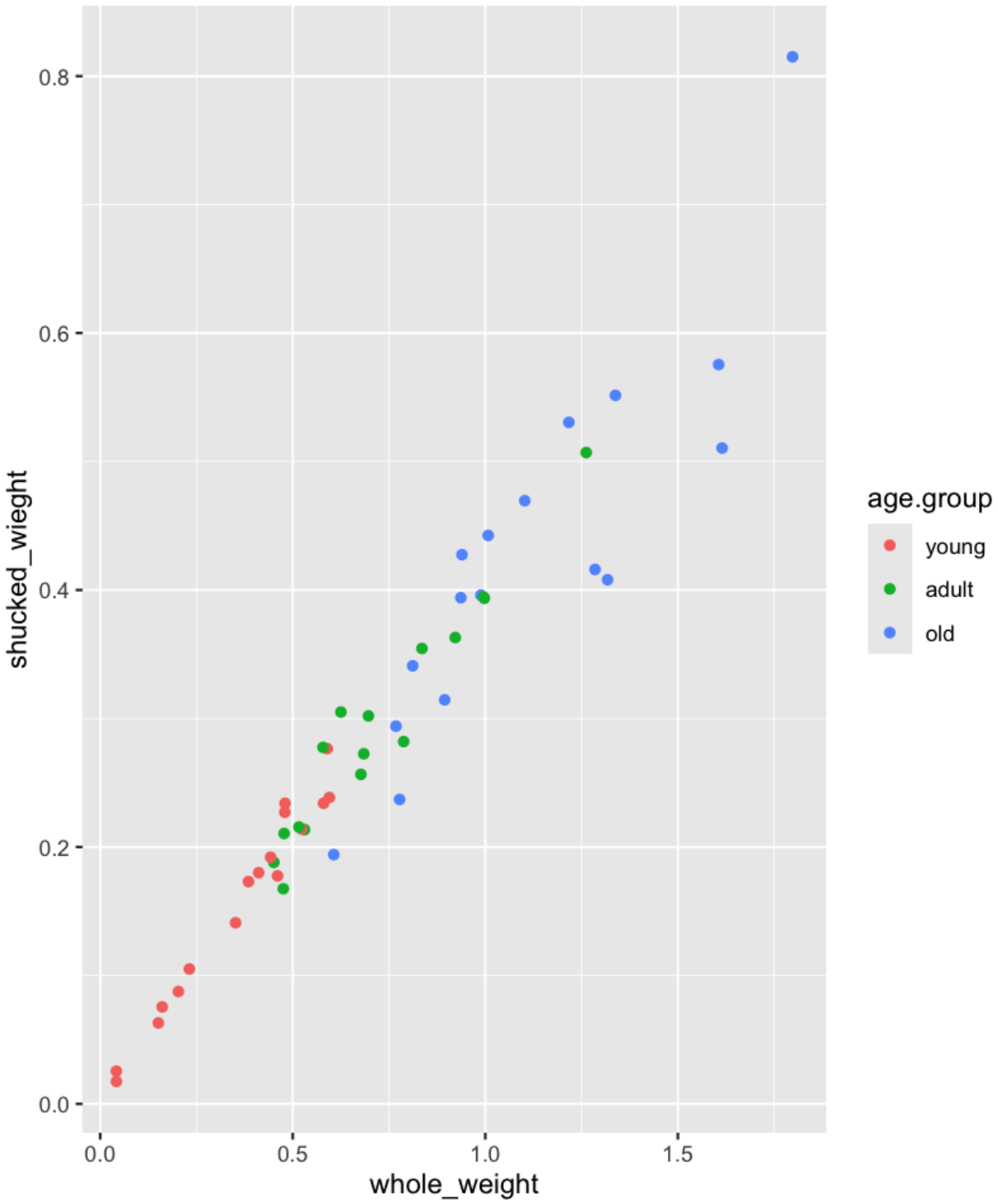
Sean McHugh
October 6, 2025
Lab 3

```
> # read dataset
> abalone.data <- read.csv("Downloads/abalone_dataset.csv")
>
> ## add new column age.group with 3 values based on the number of rings
> abalone.data$age.group <- cut(abalone.data$rings, br=c(0,8,11,35), labels = c("young",
'adult', 'old'))
>
> # creating a sample from the abalone dataset
> a.train <- sample(100,50)
>
> # create training and testing sets and run scatterplots for both models
> abalone.train <-abalone.data[a.train,]
> abalone.test <-abalone.data[-a.train,]
>
> # scatter plots for first model
> ggplot(abalone.train, aes(x = length, y = diameter, colour = sex)) +
+   geom_point()
```

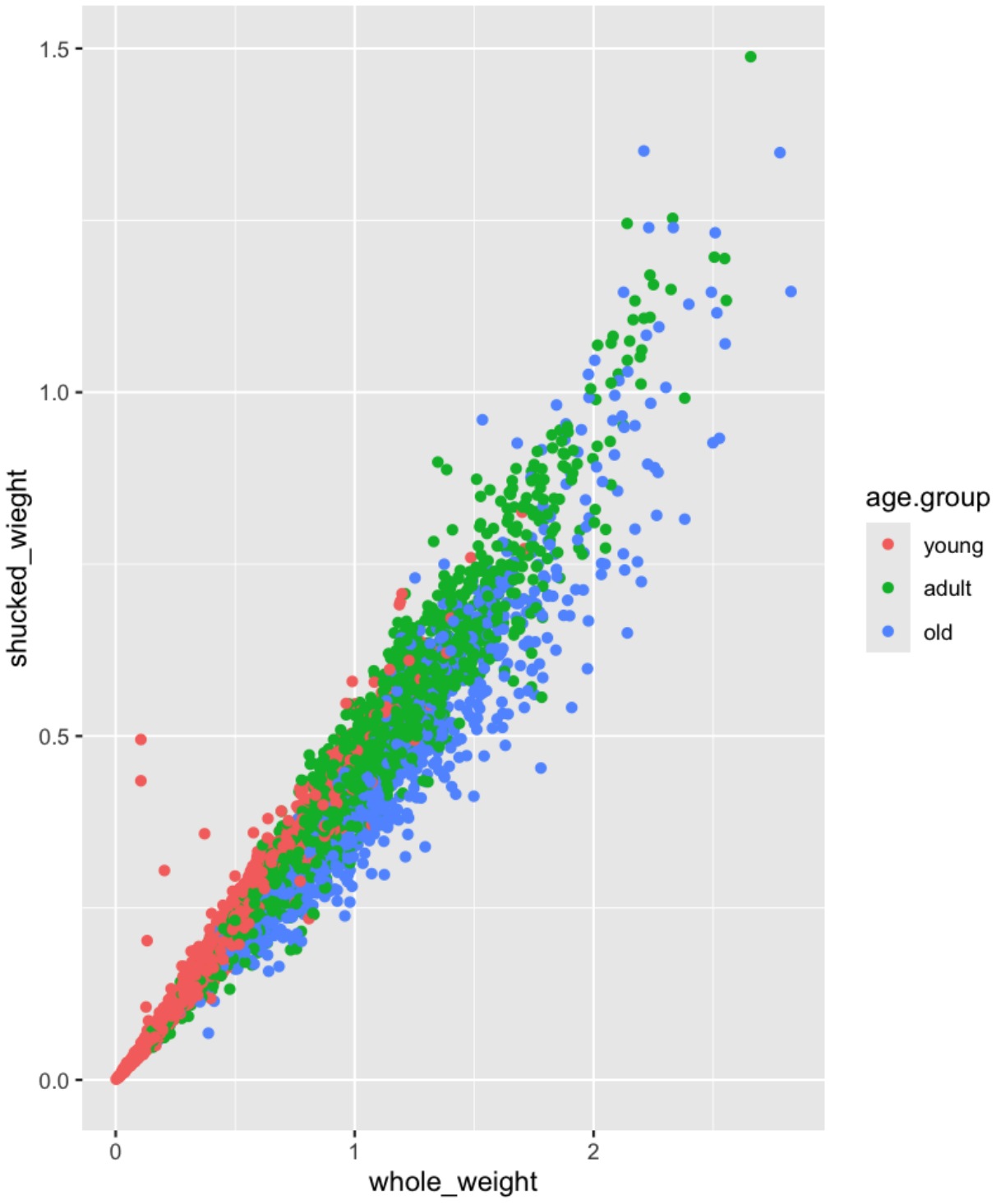



```
> ggplot(abalone.test, aes(x = length, y = diameter, colour = sex)) +
+   geom_point()
```


```
>
> # scatter plots for second model
> ggplot(abalone.train, aes(x = whole_weight, y = shucked_wieght, colour = age.group)) +
+   geom_point()
```

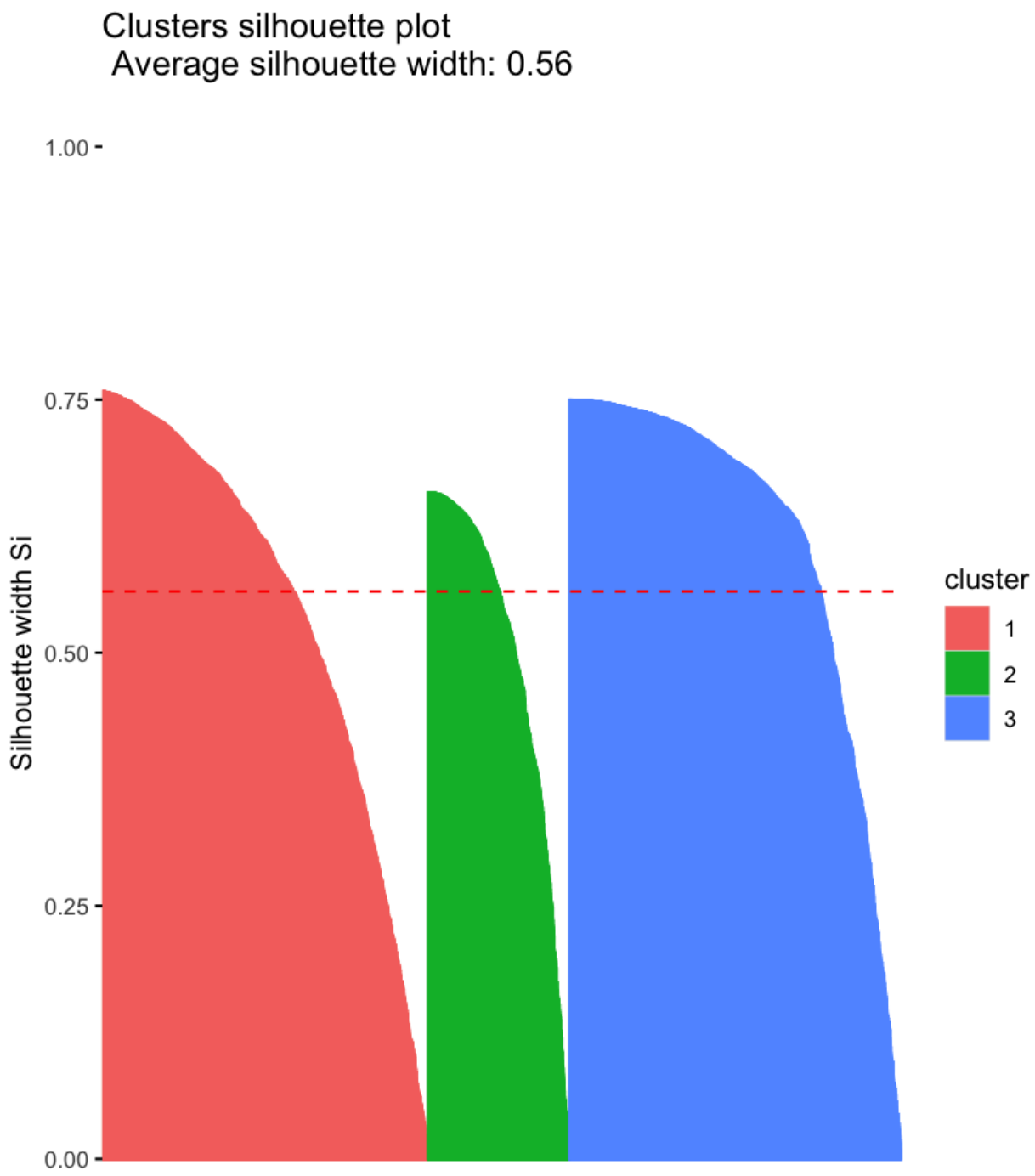



```
> ggplot(abalone.test, aes(x = whole_weight, y = shucked_wieght, colour = age.group)) +
+   geom_point()
```

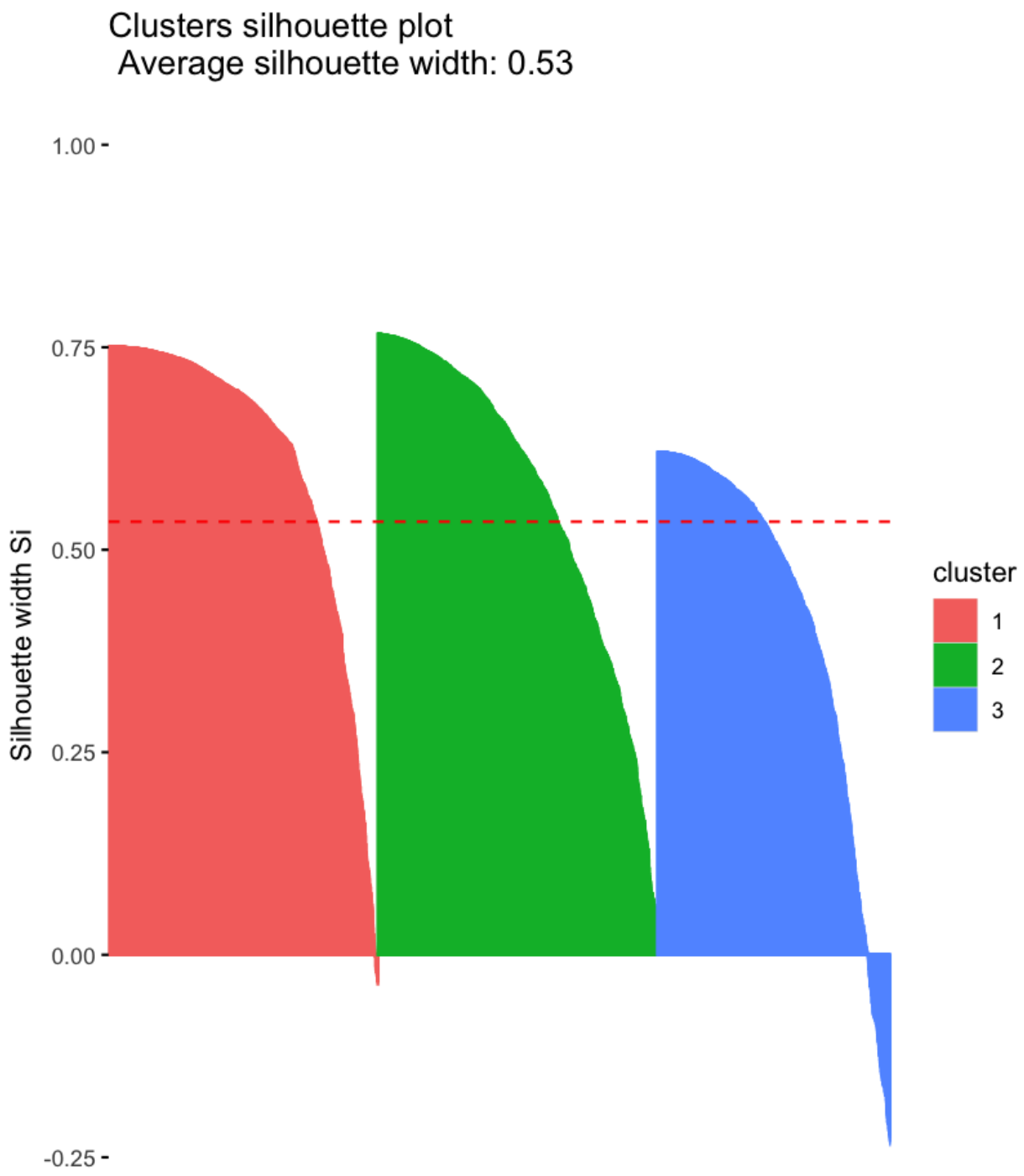



```
>
>
> ## First kNN Model and contingency table: length and diameter by sex
> knn.abalone <- knn(abalone.train[,2:4], abalone.test[,2:4], abalone.train[,10], k=3)
> table(knn.abalone, abalone.test[,10], dnn=list('predicted','actual'))
         actual
```

```
predicted young adult  old
   young   997   351  118
   adult   196   277  185
   old     197  1166  639
>
> ## Second kNN Model and contingency table: whole weight and shucked weight by age group
> knn.abalone2 <- knn(abalone.train[,5:8], abalone.test[,5:8], abalone.train[,10], k=3)
> table(knn.abalone2, abalone.test[,10], dnn=list('predicted','actual'))
         actual
predicted young adult  old
   young   971   302   95
   adult   212   315  191
   old     207  1177  656
>
> ### K-Means ###
> abalone.km <- kmeans(abalone.data[,5], centers = 3)
>
> ### Partitioning Around Medoids ###
> abalone.pam <- pam(abalone.data[,5], 3)
>
> ## Silhouette Plots
> sil <- silhouette(abalone.km$cluster, dist(abalone.data[,5]))
> sil2 <- silhouette(abalone.pam$cluster, dist(abalone.data[,5]))
> fviz_silhouette(sil)
  cluster size ave.sil.width
1       1 1703          0.54
2       2  738          0.49
3       3 1735          0.61
```




```
> fviz_silhouette(sil2)
  cluster size ave.sil.width
1       1 1438          0.61
2       2 1493          0.57
3       3 1245          0.41
```


Clusters silhouette plot
Average silhouette width: 0.53
1.00
0.75
0.50
0.25
0.00
-0.25
Silhouette width Si
cluster
1
2
3