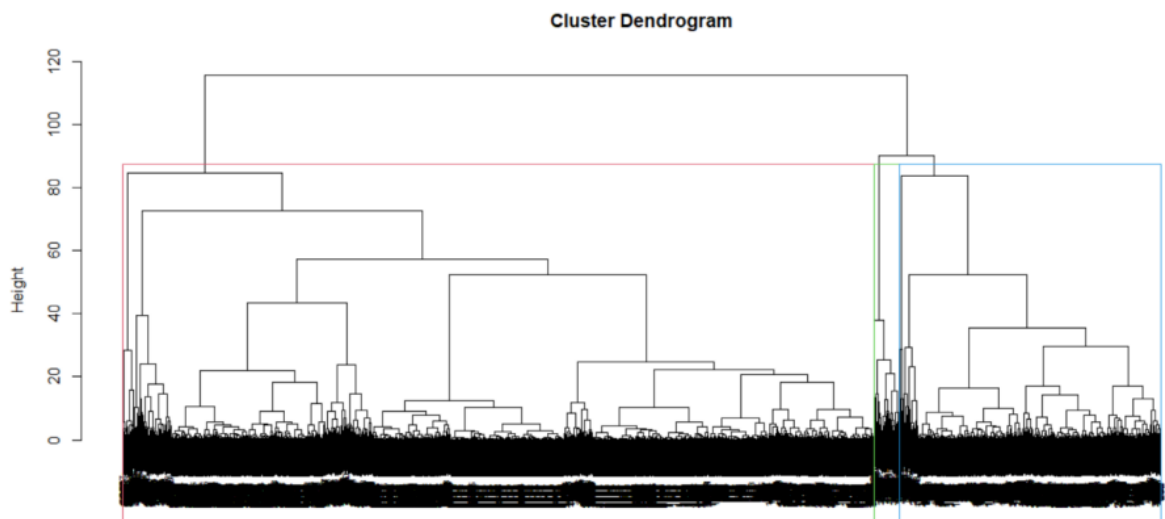


Identifying Customer segments through clustering- Hierarchical & K-Means clustering

- a) Do you need to standardize the data before applying any clustering technique? Why or why not?

Yes. It's best method to standardize the data before applying any clustering techniques as the standardization prevents variables with larger scales from dominating on how the clusters are formed and give equal weight to all. Also, whether to normalize the data or not is purely based on the objective we are trying to achieve for a specific business problem.

- b) Apply hierarchical clustering with Euclidean distance and Ward's method. How many clusters do appear?



From the plot of dendrogram, we see that as there are larger numbers of observations and we see many datapoints are cluttered at the lower height region of 0-20. Also, as it's hard to interpret those many numbers of clusters we go to higher level of height on Y-axis, say above 80 to get fair number of clusters. Also, because our business problem is to identify the categories of customers to target for mileage offers it's ideal to go for 3 clusters that have frequent, not so frequent & normal travellers.

- c) Compare cluster centroids to characterize different clusters and try to give each cluster a label—a meaningful name that characterizes the cluster.

```
> Airlines_final_summary
```

	[,1]	[,2]	[,3]
Cluster	1.00	2.00	3.00
Cluster_count	2893.00	1010.00	96.00
Balance	-0.23	0.58	0.74
Qual_miles	0.04	-0.14	0.32
cc1_miles	-0.48	1.34	0.44
cc2_miles	0.04	-0.10	-0.10
cc3_miles	-0.06	0.19	-0.06
Bonus_miles	-0.42	1.11	1.05
Bonus_trans	-0.32	0.72	2.18
Flight_miles_12mo	-0.09	-0.15	4.44
Flight_trans_12	-0.10	-0.15	4.53
Days_since_enroll	-0.21	0.57	0.45

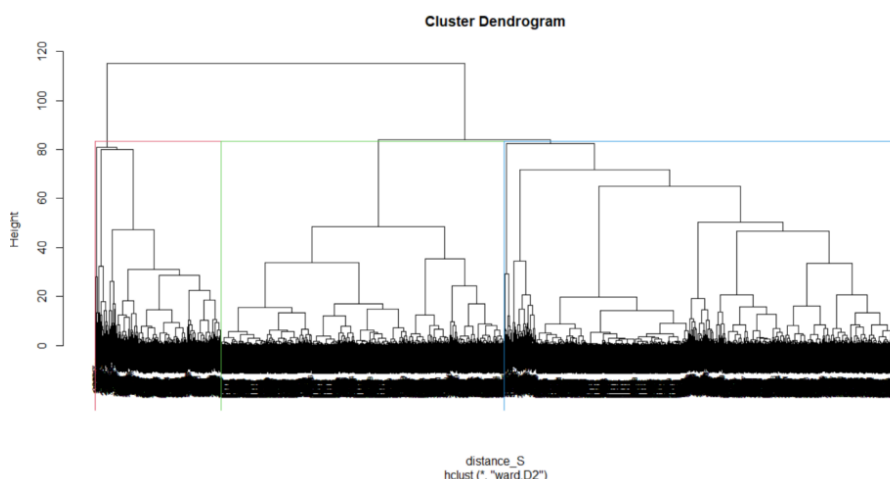
From the table obtained in 'Airlines-final summary', we can assess the following and name the cluster accordingly:

#Cluster-1: This cluster comprises of 72% of total customers. This segment has recently joined customers due to which there is very a smaller number of balance accrual, miles travelled, and reward earned in the past 12 months. This group can be categorized as "New Travelers"

#Cluster-2: This cluster comprises of 25% of total customers. This segment of customers are oldest customers of the airlines and have decent number of Bonus miles and balance accrual. But the flight miles travelled in the past 12 months is comparatively less which tells us that this group of customers are not very frequent travellers. This cluster can be called as "Loyal, but not very frequent travellers".

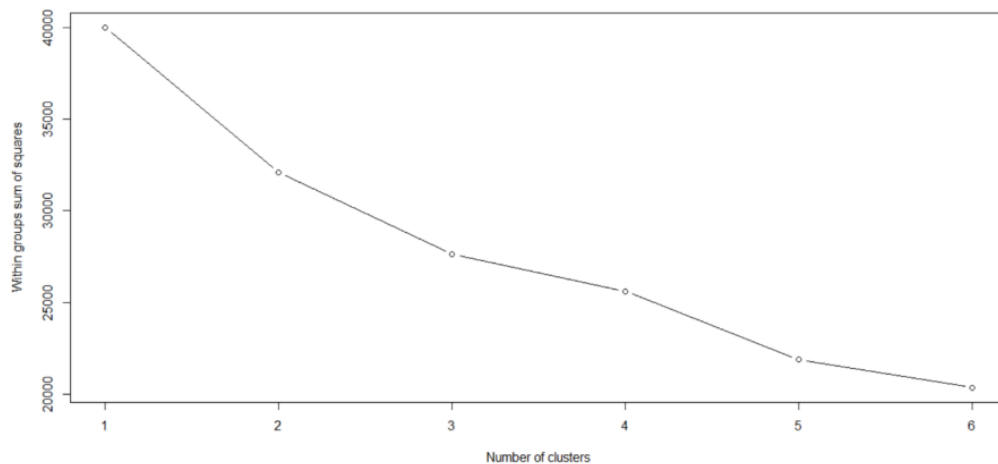
#Cluster-3: This cluster comprises of least number of total customers that is 2% which are new. These customers covered the highest number flight miles in the past 12 months and topped in flight transactions due to which there is good amount of bonus miles accrual. Customers in this segment are next top in oldest customers to the airlines. Hence this cluster can be called as "frequent travellers".

- d) To check the stability of clusters, remove a random 5% of the data (by taking a random sample of 95% of the records), and repeat the analysis. Does the same picture emerge?



When considering only 95% of the sample data to check the stability of the clusters, it is observed that we are reaching at same number of clusters when we cut the dendrogram obtained from the sample 95% data. With both complete data & sample data, we had to cut the tree at the height of between 85-88 (shown above in the image) to obtain 3 clusters.

- e) Cluster all passengers again using k-means clustering. How many clusters do you want to go with? How did you decide on the number of clusters? Explain your choice on the number of clusters.



In K-Means clustering, chose the random center point of 3 based on our business need & ran the elbow curve to find the maximum number of clusters that can be taken. The elbow curve (image below) showed a shift in slope at $k=2$ and slightly going up after $k=3$. Since 2 clusters won't be helpful in our analysis and up to 3 clusters is a optimum number based on the curve, it's decided to go with 3 clusters.

Also, the 'eclust' model of 'factoextra' gave $K=3$ as optimum number of clusters. Going $K>3$ is giving over-lapping imagery which indicates cluttered clusters. Hence $K=3$ is optimal number.



- f) How do the characteristics of the clusters, obtained in Part (e), contrast or validate the finding in Part c above?

Comparing the results of both Hierarchical & k-means clustering:

- i) Both the clustering techniques have given close to same results in majority of parameters.
- ii) It is observed that when it comes to miles earned through 'Reward's credit card' and 'Business credit Card', the results differed in case of frequent flyers category. K-means clustering failed to show how frequent fliers would have high credits earned through frequent flyer card in CC1_miles whereas, Hierarchical clustering captured it accurately.

Overall, we can say that Hierarchical clustering gave better results compared to Kmeans in different categories bases on above analysis.

- g) Which cluster(s) would you target for offers, and what type of offers would you target to customers in that cluster? Include proper reasoning in support of your choice of the cluster(s) and the corresponding offer(s).

Following are recommendations to the East-West Airlines regarding offers for the 3 categories of customers:

- i) **Frequent travellers:** As this category comprises of low percentage of overall customers and is doing good at the current plan, it's advisable for the management to keep up with same offers to this category.
- ii) **New travellers:** As the customers belonging to this category are recent new travellers and comprise of large majority of customers, management can offer special discounts to convert them into frequent travellers which in turn can increase the reward points & frequent traveller points to motivate them to choose this airline for their next travel.
- iii) **Loyal, but not frequent travellers:** As this category of customers already have good points pertaining to bonus miles, management can try to retain these customers by providing reward points which can be used next travel bookings which in turn can increase frequent traveller reward points.