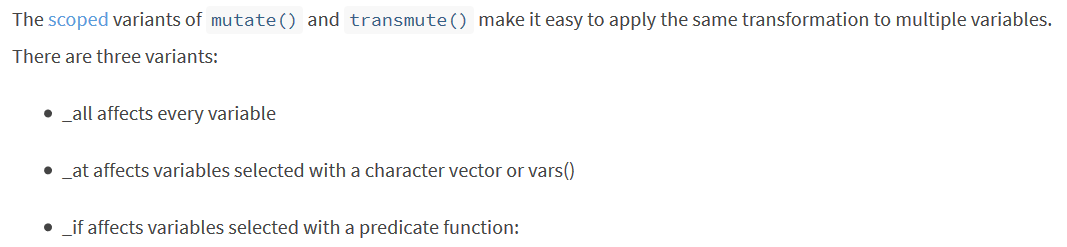
Degrees that pay you back

In this project I learned:-

1. Sometimes data containing Salary figures is stored as String because of the presence of $ before the figure. To deal with such a problem, we can use gsub() to identify the sign in the string then replace it with space and convert all of it into numeric. We can do this for values in different columns using mutate\_at().
2. 
3. We use Elbow Method, Silhouette Method and Gap- Statistic Method to determine the optimum number of clusters for k-means algorithm.
4. We use fviz\_nbclust() which determines and visualize the optimal number of clusters using different methods: within cluster sums of squares, average silhouette and gap statistics. For Elbow Method we use method=”wss”
5. Instead of needing to "manually" apply the elbow method by running multiple k\_means models and plotting the calculated the total within cluster sum of squares for each potential value of k, fviz\_nbclust handled all of this for us behind the scenes. We use it for silhouette method as well.
6. The Silhouette Method will evaluate the quality of clusters by how well each point fits within a cluster, maximizing average "silhouette" width.
7. The Gap Statistic Method will compare the total variation within clusters for different values of k to the null hypothesis, maximizing the "gap.". We have the clusGap function to calculate this behind the scenes and the fviz\_gap\_stat function to visualize the results.
8. Through 3 different methods, we find 3 is appropriate number of clusters. Now, we run the kmeans algorithm and find the clusters. Then we mutate the cluster assignment to original data and visualize the classification for features of our interest through a graph.
9. Now we wanted to plot a graph for each of the three clusters where we wanted to see Salary and Percentile for different majors. With the current data this was tough to do, so we used gather() for Data Manipulation in such a way that we get the data in “Long Format”.
10. At last, we plotted graphs for 3 clusters – Majors with low Salary , Majors with Medium Salary and Majors with High Amount of Salary.