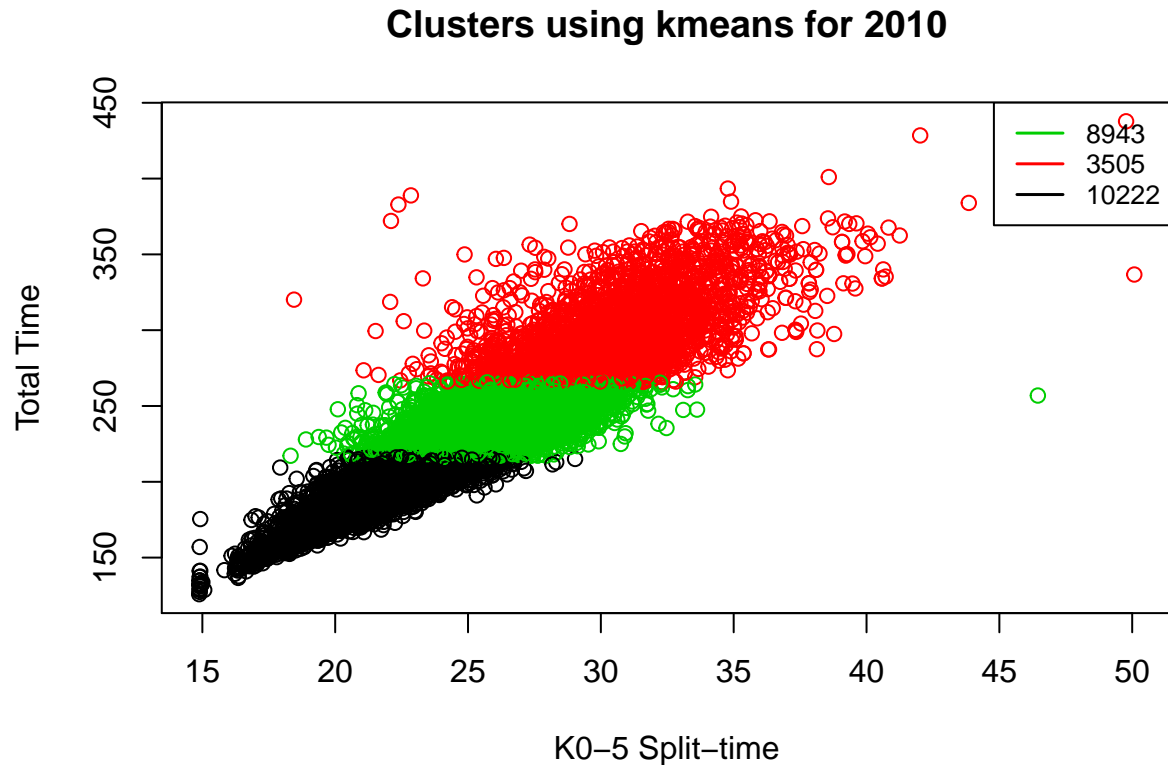


Clustering Analysis

K-means clustering is a simple method for partitioning a dataset into K distinct clusters. In order to find clusters between total time for running a marathon and time to run first five kilometers, we used R to generate clusters using kmeans function. We set the number of clusters to be k=3 inside the kmeans function. We created a data frame that contains total time and time to run first five kilometers for 2010 and passed it as a parameter to kmeans function. We plotted the clusters as there are only two variables.

```
kc <- kmeans(df, k=3, nstart=20) plot(df, col=kc$cluster, xlab="K0-5 Split-time", ylab="Total Time",  
main="Clusters using kmeans")
```



For further clustering analysis for more than two variables, we would perform PCA and plot the first two principal components score. vectors.