K-Means

Cody Frisby

9/30/2017

Using 20 observations, I produce the plot using the R API of `plotly`. Cool thing about rendering a scatterplot using `plotly` is that it is interactive. It appears that there are 3 or 4 groups depending on how to "slice" it. I can see a group of the older individuals, all grandparents. I can see one of the short/lower-weight individuals (all children) as well. Centroids could be

$$Centroid_1 = (age = 10,\ weight = 57,\ height = 53)$$
$$Centroid_2 = (age = 30,\ weight = 150,\ height = 69)$$
$$Centroid_3 = (age = 66,\ weight = 180,\ height = 67)$$

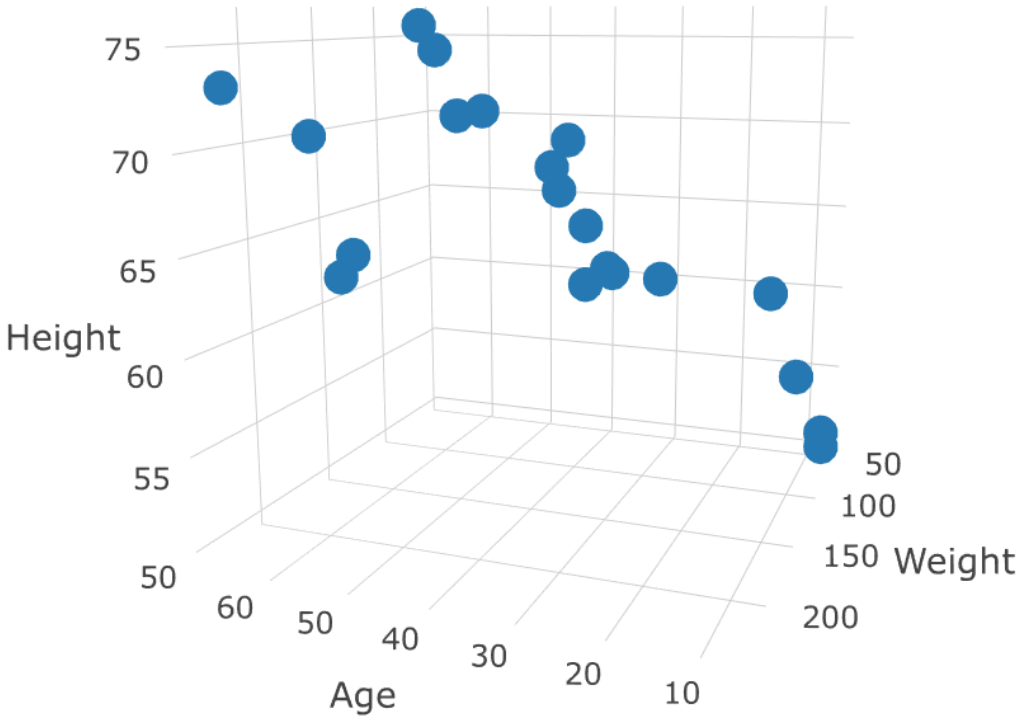

Figure 1: 3D Scatterplot

If we choose $k = 3$ using k-means clustering and then coloring the points by group we can see how well we do.

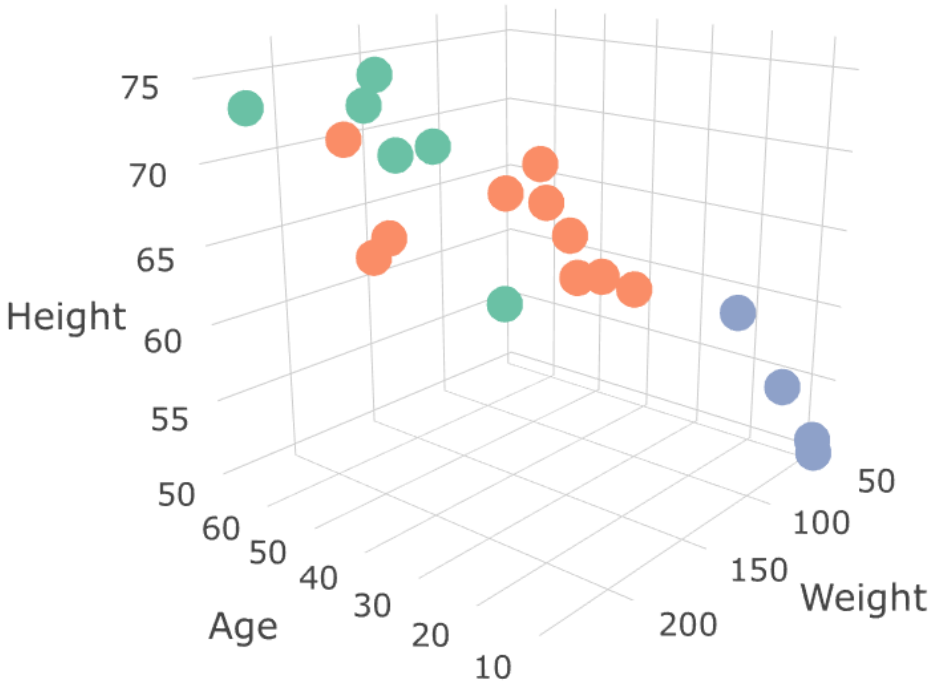

Figure 2: 3D Scatterplot with Kmeans Groups

Not too bad, if we orient it right we can see the distinct groups better. *Kmeans* includes the taller/heavier oldest individual in the tall/heavier group where I would have grouped that individual in the "grandparents" group. But other than that it doesn't look too bad. For reference I show the kmeans centroids.

height	weight	age
52.00000	59.0000	9.00000
72.00000	215.0000	42.66667
66.80000	161.0000	53.20000
63.33333	123.3333	29.33333

What if we sliced it using 3 groups and hiearchal clustering with single linkage, still using $k = 3$ as the number of clusters.

Groups are different, I'd even argue they are " better".