Hierarchical Clustering Pada Menu Minuman McDonalds

18523202/18523223/18523233

Data Menu Makanan Pada analisis data ini kami menggunakan sebuah dataset dari Kaggle dataset yang telah dimodifikasi dengan memilih menu minuman dan memangkas beberapa fitur tanpa mengubah originalitas data yang ada, sehingga menjadi dataset baru bernama `menuminuman`, dengan dataset asli bernama (`Facts for McDonald's Menu`). Data ini berisi beberapa kandungan makanan dan minuman yang disajikan oleh restoran cepat saji McDonalds beserta kecukupan hariannya, dari setiap minuman tersebut mmemiliki kandungan seperti jumlah porsi, kalori, gula, dan karbohidrat. Rinciannya adalah : kategori minumannya (`Category`), minuman yang disajikan (`item`), ukuran sajian setiap porsi (`Serving.Size`), kandungan kalori (`Calories`), Karbohidrat (`Carbohydrates`), Persentase kecukupan harian karbohidrat(`Carbohydrates % Daily.Value.`), serta kandungan gulanya (`Sugars`).

Berikut disajikan dataset `menuminuman`.

```
menuminuman = read.csv('menuminuman.csv')
summary(menuminuman)
```

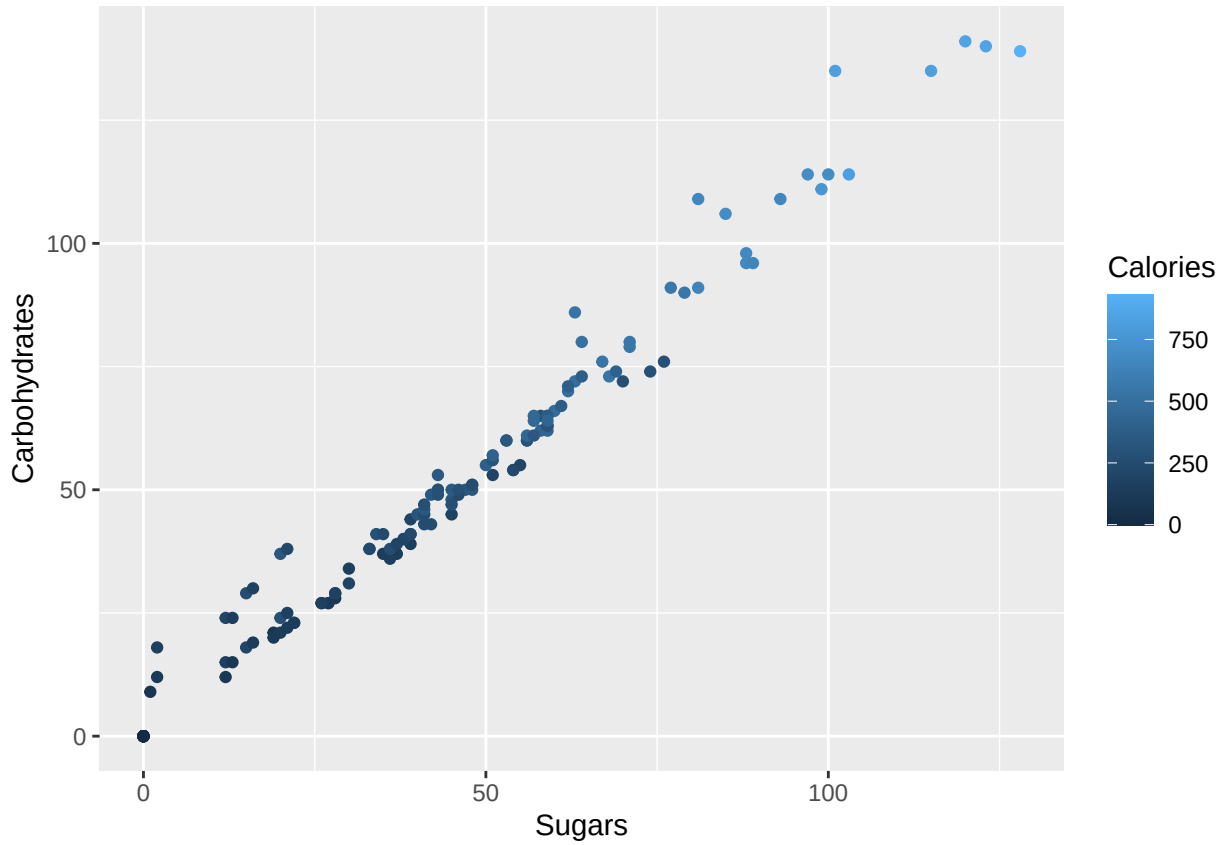
```
##    Category             Item           Serving.Size          Calories
##  Length:141         Length:141         Length:141         Min.   :  0.0
##  Class :character   Class :character   Class :character   1st Qu.:150.0
##  Mode  :character   Mode  :character   Mode  :character   Median :270.0
##                                                           Mean   :301.4
##                                                           3rd Qu.:410.0
##                                                           Max.   :930.0
##  Carbohydrates    Carbohydrates....Daily.Value.     Sugars
##  Min.   :  0.00   Min.   : 0.00                 Min.   :  0.00
##  1st Qu.: 27.00   1st Qu.: 9.00                 1st Qu.: 21.00
##  Median : 47.00   Median :16.00                 Median : 43.00
##  Mean   : 49.51   Mean   :16.51                 Mean   : 43.87
##  3rd Qu.: 65.00   3rd Qu.:22.00                 3rd Qu.: 59.00
##  Max.   :141.00   Max.   :47.00                 Max.   :128.00
```

Proses Analisis Data (Exploratory Data Analysis) Sistem pembuatan grafik dengan 'ggplot' dapat dilakukan dengan menggunakan 'ggplot2' yang merupakan implementasi dari konsep *Grammar of graphic* untuk bahasa pemrograman R.Membuat analisis dari dataset 'menuminuman' yang tersedia dalam paket 'ggplot2'.

```
library(ggplot2)
```

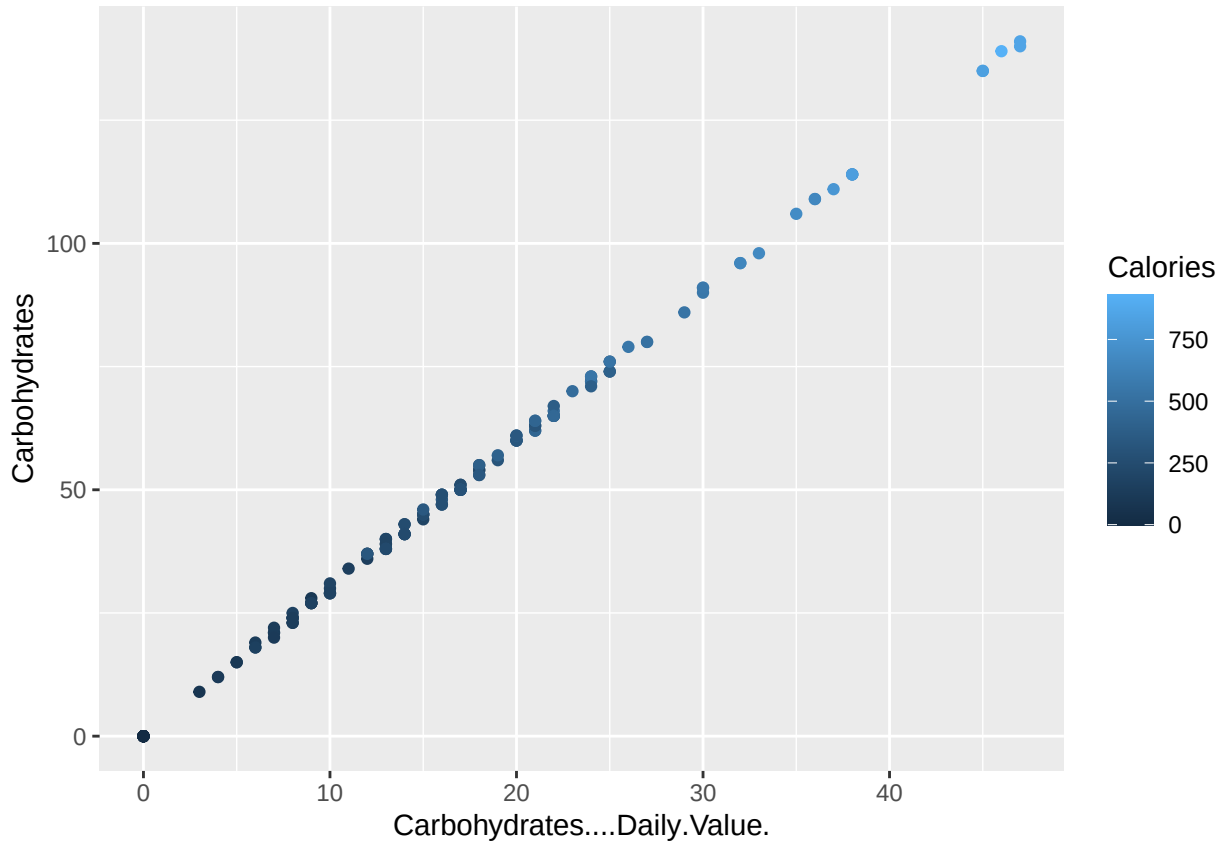
Selanjutnya kami menggunakan fungsi 'qplot()' untuk membuat grafik menggunakan 'ggplot2'.

```
qplot(x = Sugars, y = Carbohydrates, colour = Calories, data = menuminuman)
```



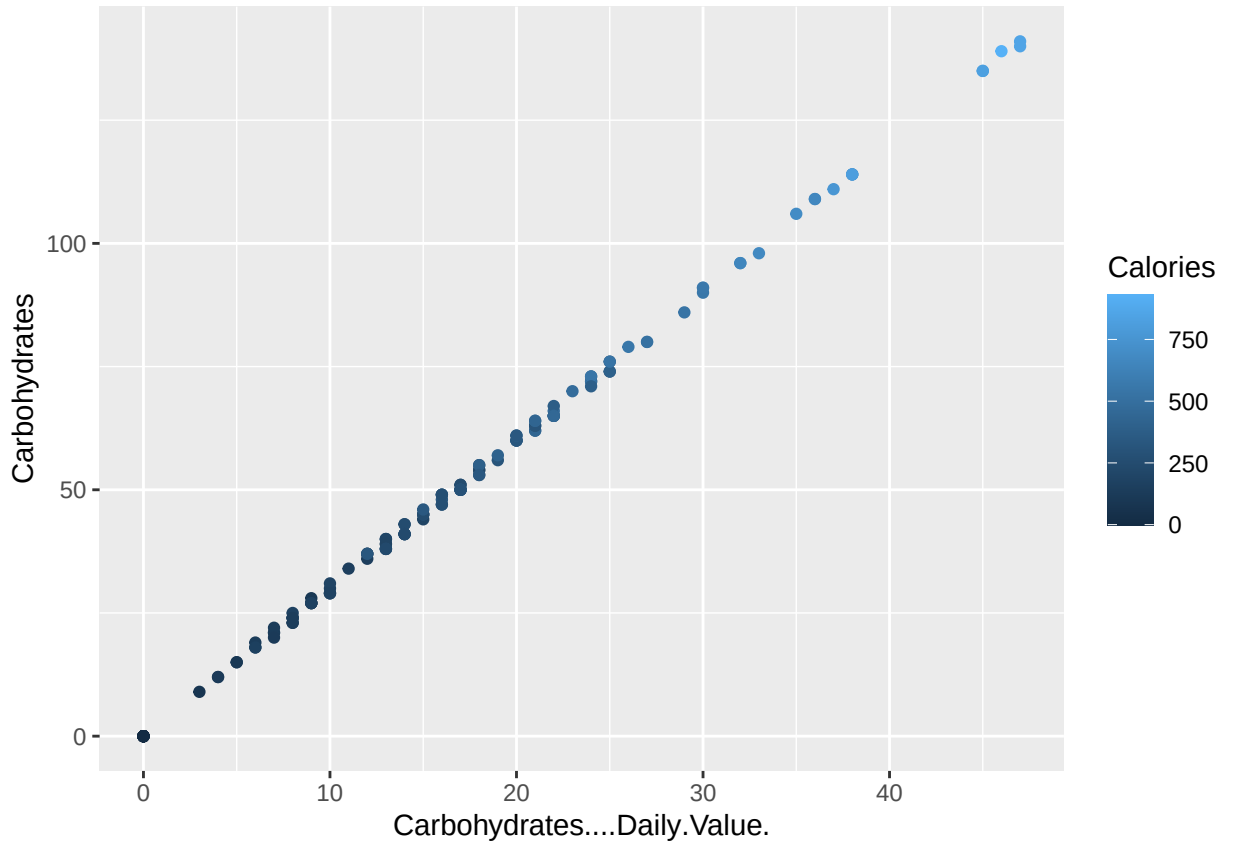
Agar visualisasi data dapat dilihat dengan lebih leluasa, maka grafik di atas dapat diolah dengan menggunakan penulisan kode sebagai berikut:

```
ggplot(data = menuminuman,
       mapping = aes(x = Carbohydrates....Daily.Value., y = Carbohydrates, colour = Calories)) +
  geom_point()
```



Berikut untuk menyimpan grafik ke dalam boyek R bernama 'plot_menuminuman' dan kemudian menyimpannya dalam file komputer dalam satu folder dengan project .Rmd yang sedang dijalankan dengan nama berkas 'minuman.png.

```
plot_minuman <- ggplot(data=menuminuman) +
  geom_point(mapping = aes(x = Carbohydrates....Daily.Value.,
                           y = Carbohydrates, colour = Calories))
plot_minuman
```




```
ggsave(filename = "minuman.png", plot = plot_minuman)
```

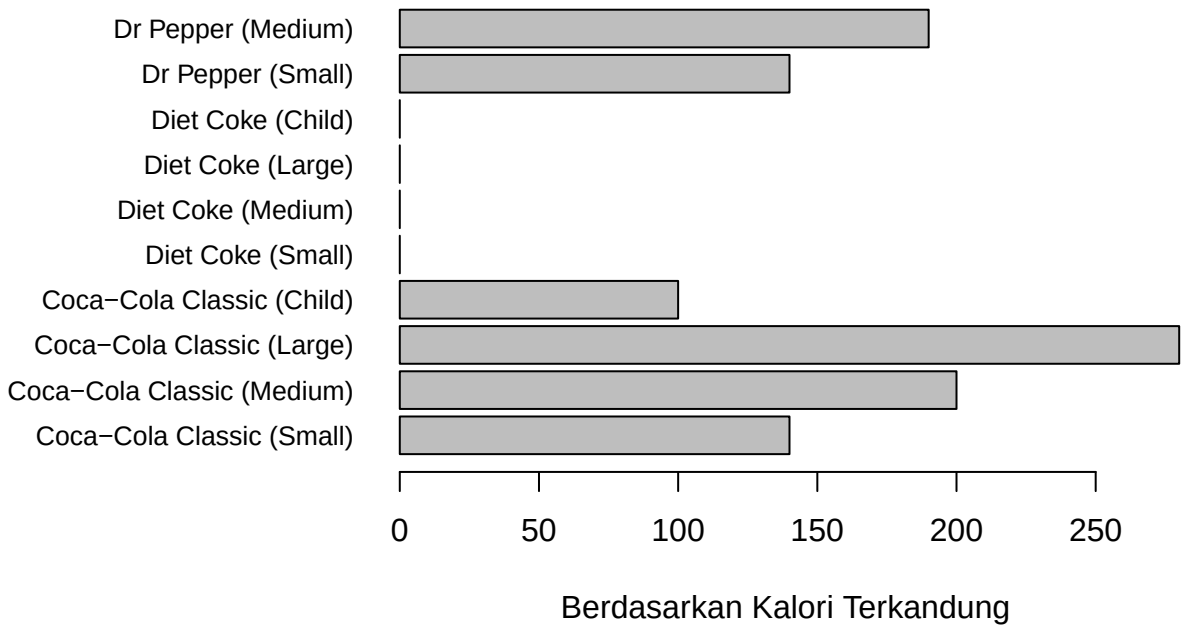
```
## Saving 6.5 x 4.5 in image
```

Dari sebaran setiap variabel berdasarakan empat barplot di atas, kami mendapatkan informasi tentang data menuminuman tersebut bahwa fitur Carbohydrates, Sugars, dan Calories memiliki nilai yang sebanding.Menurut data pada fitur Carbohydrates. . . .Daily.Value menunjukan persentase karbohidrat yang terkandung cukup tinggi, artinya beberapa minuman memang sangat tinggi gula dan ada juga yang sangat rendah bahkan tidak ada sama sekali.

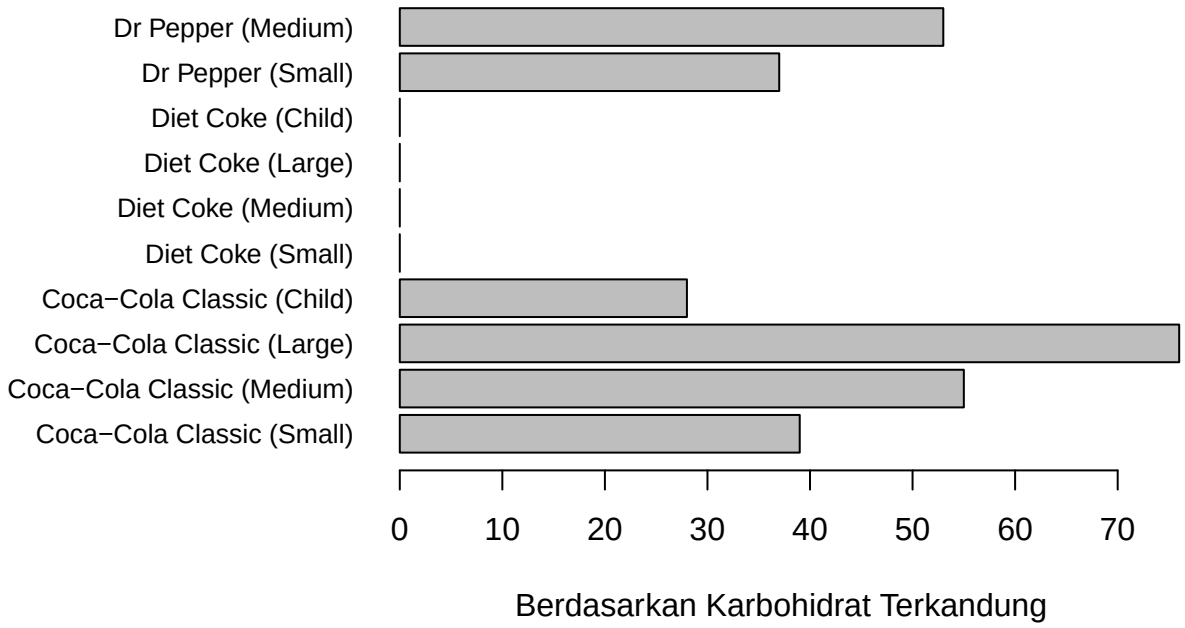
Kemudian kami menggunakan sebuah alternatif untuk mendapatkan visualisasi dari setiap variabel menggunakan `barplot()`. Pada baris pertama kami membuat sebuat data frame dari data set `menuminuman` dan menambahkan satu kolom berisi kategori menu minuman (Category). Pada baris kedua kami tampilkan nama menu minuman yang ingin di visualisasikan, untuk mengubah orientasi label pada sumbu vertikal sehingga dapat terbaca secara horizontal (Item) kebetulan data nama minuman yang kami tampilkan data minuman nomor 1 hingga 10(random bebas menampilkan nomor berapa), lalu Baris ketiga dan keempat untuk mengatur margin agar sesuai dengan yang kami harapkan, pada baris ini kami atur agar visualisasi tepat menampilkan nama minuman (bersifat random sesuai dengan keinginan letak dari data yang di visualisasi). Baris kelima berfungsi untuk membuat barplot untuk variabel kalori `Calories`; kemudian karbohidrat`Carbohydrates`; lalu persentase karbohidrat harian dari minuman`Carbohydrates....Daily.Value`; serta kandungan gula yang ada didalam minuman tersebut`Sugars`.

```
df <- data.frame(Category=rownames(menuminuman), menuminuman)
df <- df[1:10 ,]
par(las=1)
```

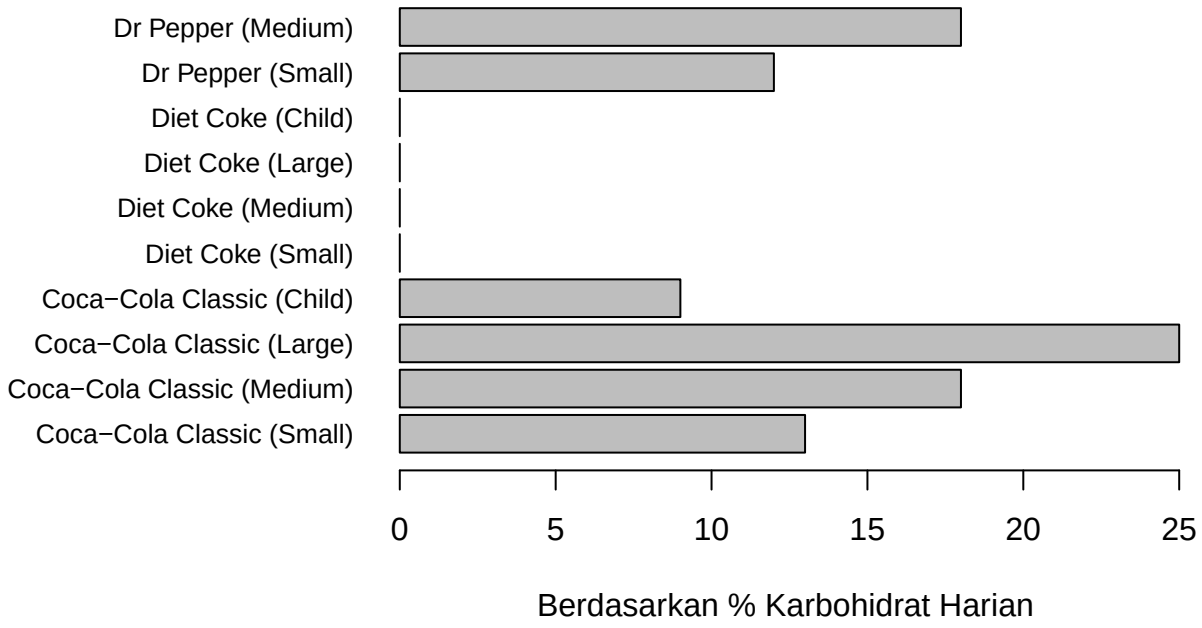
```r
par(mar=c(4,10,6,2))
barplot(df$Calories, names.arg = df$Item, horiz = TRUE, cex.names = 0.8,
        xlab = "Berdasarkan Kalori Terkandung")
```



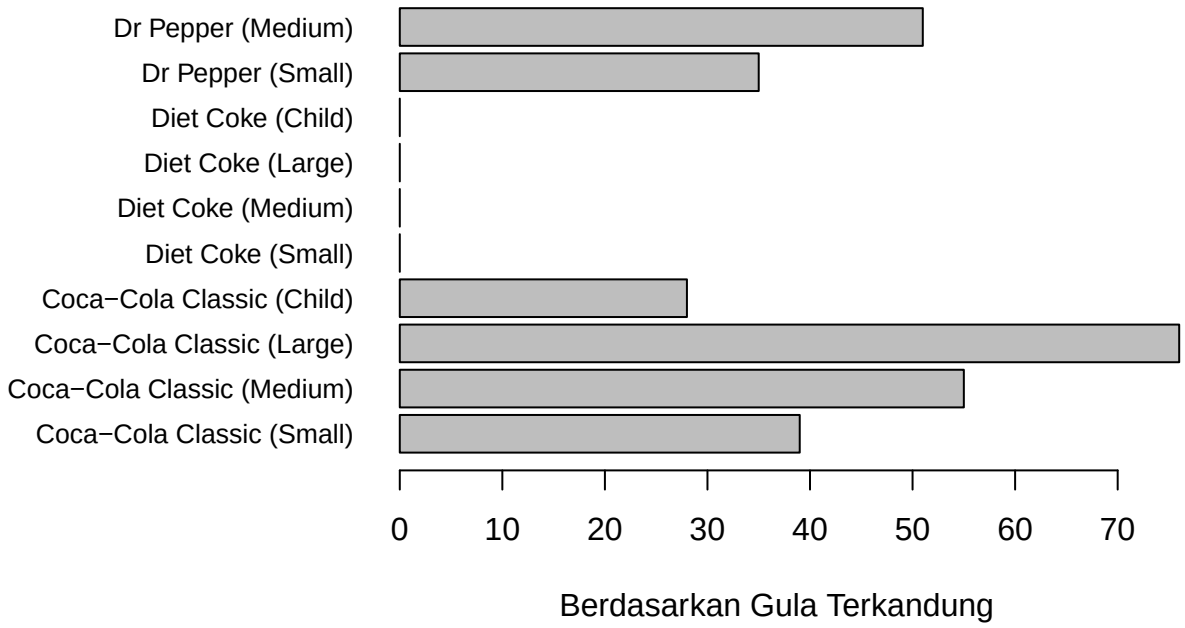

```r
barplot(df$Carbohydrates, names.arg = df$Item, horiz = TRUE, cex.names = 0.8,
        xlab = "Berdasarkan Karbohidrat Terkandung ")
```




```r
barplot(df$Carbohydrates....Daily.Value., names.arg = df$Item, horiz = TRUE, cex.names = 0.8,
        xlab = "Berdasarkan % Karbohidrat Harian")
```

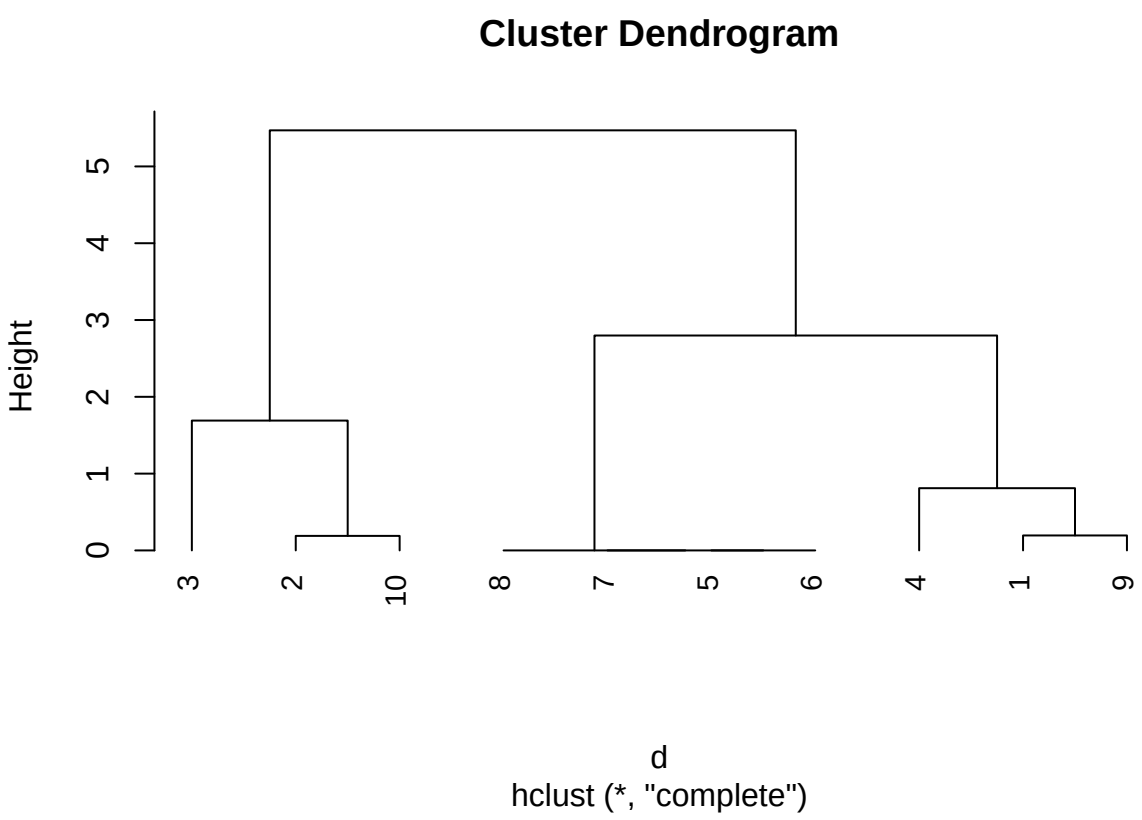



```r
barplot(df$Sugars, names.arg = df$Item, horiz = TRUE, cex.names = 0.8,
        xlab = "Berdasarkan Gula Terkandung  ")
```



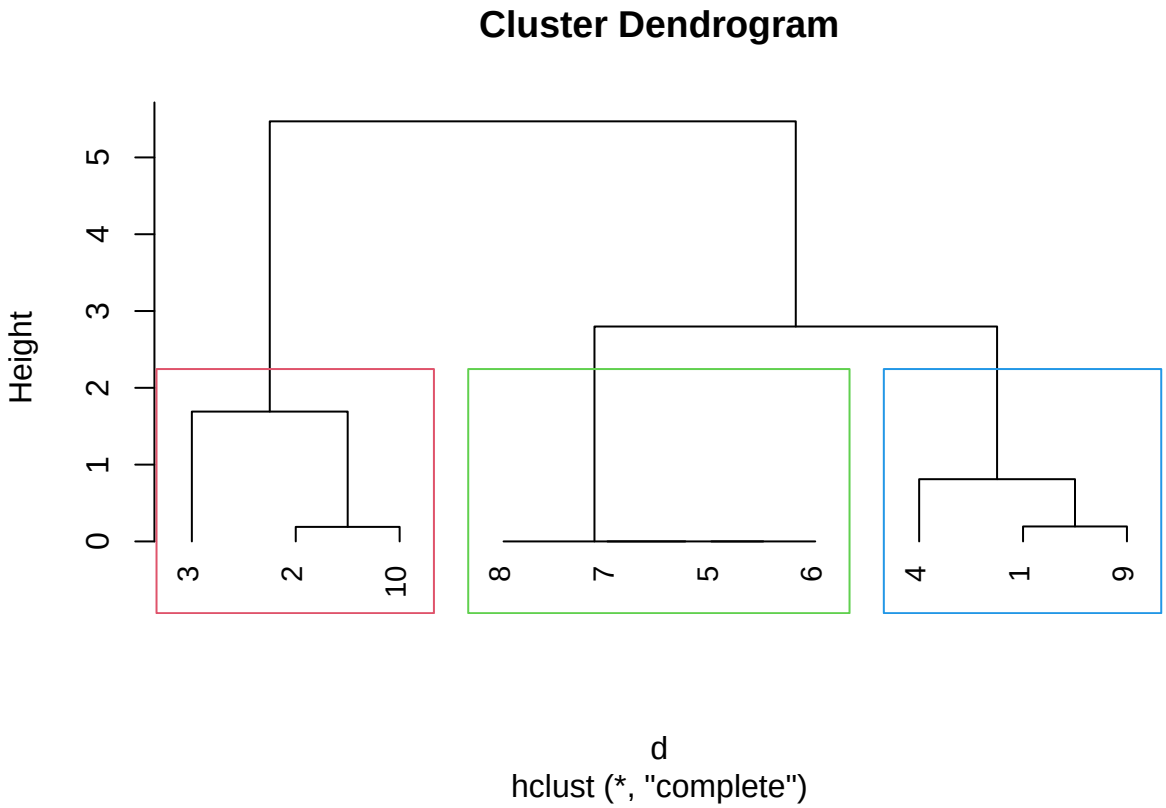
Hierarchical Clustering Setelah melakukan analisis di sini kami mencoba menggunakan metode hierarchical clustering menggunakan fungsi `hclust`. Pada bagian ini hanya digunakan empat variabel, maka variabel `Item` dihilangkan menggunakan source code Baris 1. Fungsi `hclust` menerima matriks jarak (dissimilarity meassure dari setiap pasang variabel), maka kami menghitung matriks tersebut mengunakan Baris ke 2. Pada Baris 3, kami melakukan hiearchical clustering dengan metode `complete linkage` Dengan memperlakukan data sebagai kelompok, selanjutnya kami pilih jarak dua kelompok yang terkecil.Pada baris ke 4 membuat plot dendrogram dari hasil clustering; parameter `cex` emngatur besar font untuk label pada sumbu x, `hang` mengatur posisi label terhadap sumbu y.

```
1 df <- scale(df[, c(5,6,7,8)])
2 d <- dist(df, method = "euclidean")
3 clusters <- hclust(d, method = "complete" )
4 plot(clusters, cex = 0.9, hang = -1)
```



Kemudian kami menggunakan fungsi `rect.hclust()` untuk menggambar kotak pada sejumlah klaster yang kami inginkan agar klaster yang terbentuk dapat jelas terlihat. Sebagai contoh, kami ingin melihat 3 klaster dari hasil clustering di atas. Maka kami set `k=3`(bersifat random sesuai dengan klaster yang diharapkan asalkan klaster lebih sedikit dari jumlah data) dan terdapat parameter `border` mengatur warna kotak dari setiap klaster.Klaster pertama untuk minuman yang memiliki kandungan gula tinggi dengan kotak warna merah, klaster kedua untuk minuman yang tidak ada kandungan gula sama sekali, dan klaster ketiga untuk minuman yang memiliki kandungan gula tidak terlalu tinggi.

```
plot(clusters, cex = 0.9, hang = -1)
rect.hclust(clusters, k = 3, border = 2:5)
```



Dapat ditarik kesimpulan bahwa kandungan beberapa minuman yang terdapat di restoran cepat saji Mcdonalds memiliki kandungan karbohidrat, kalori dan gula yang cukup tinggi, namun ada beberapa minuman juga yang rendah karbohidrat, kalori dan gula bahkan ada yang sama sekali tidak ada. Untuk struktur penulisan kode R di atas setidaknya terdapat tiga komponen utama untuk membuat grafik, yaitu :

1. *Data*
2. *Aesthetic mapping*
3. *Geometric object*

Bentuk *Geometric object* dalam kode berikut :

```
?aes
```

```
## starting httpd help server ... done
```

```
?geom_point
```