

# CS5542 Big Data Apps and Analytics

## LAB ASSIGNMENT #2

### REPORT and SCREEN SHOTS

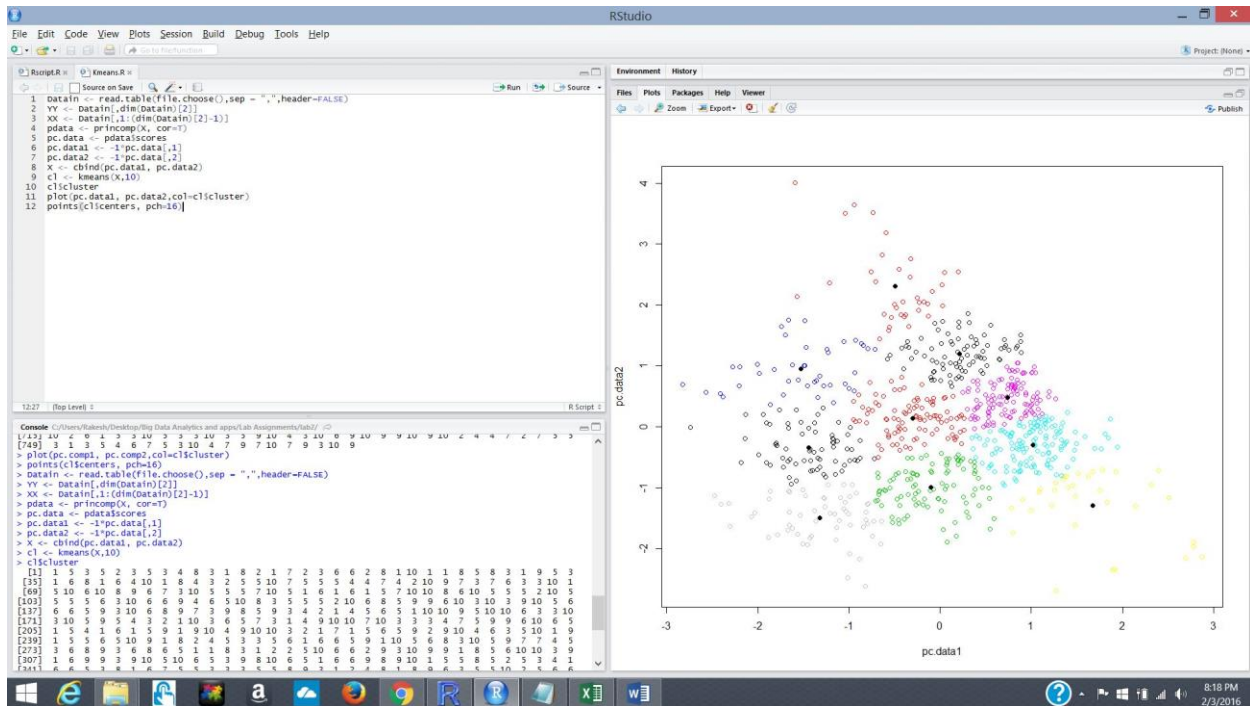
#### 1. R Project

**Prepare a dataset and perform k-means clustering.**

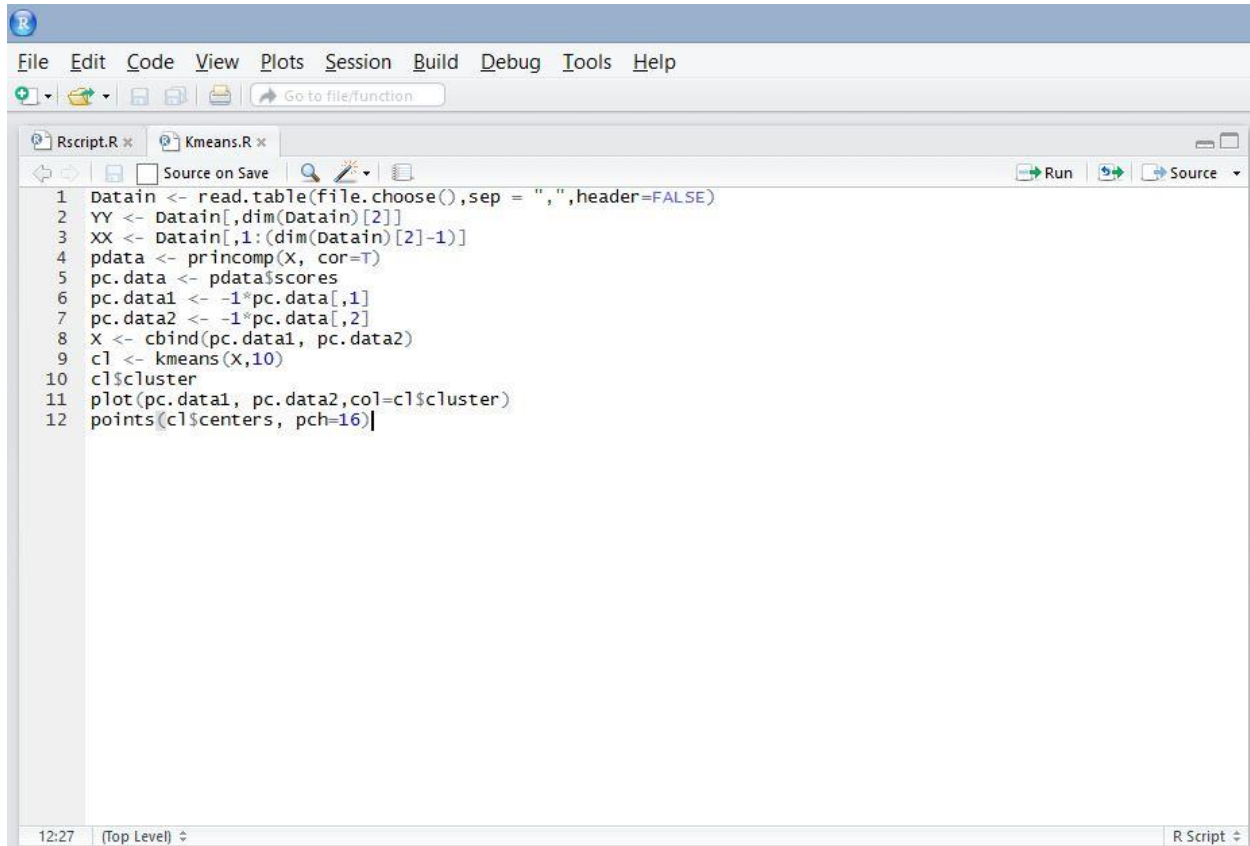
The diabetes data set is taken from the UCI machine learning database repository at:

<http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>

1. We LOAD the data into R using the file.choose() method and read csv file into Dataframe in R.



## 2. Write the R script in R studio.



The screenshot shows the R Studio interface with a script editor containing the following R code:

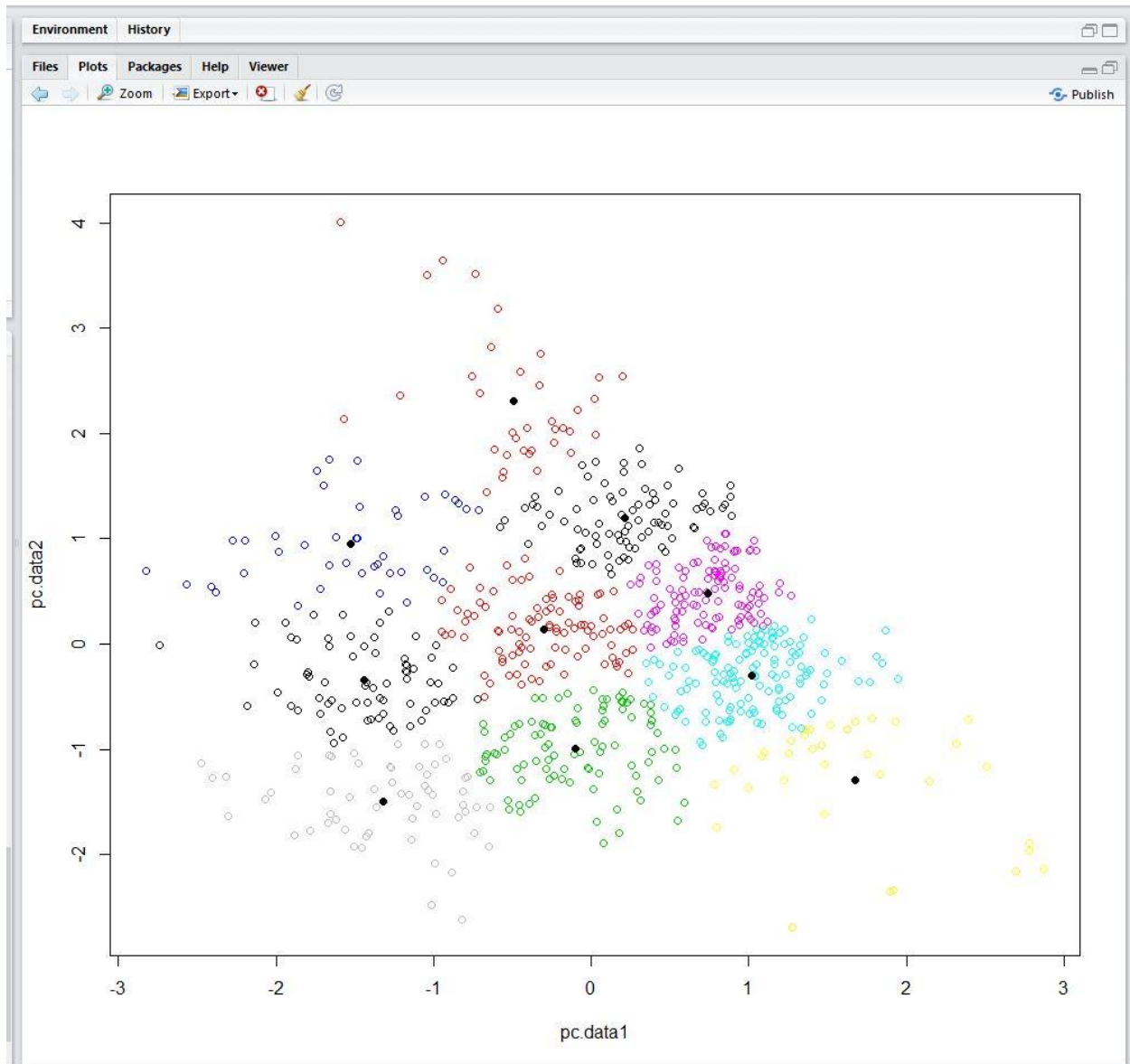
```
1 Datain <- read.table(file.choose(), sep = ",", header=FALSE)
2 YY <- Datain[, dim(Datain)[2]]
3 XX <- Datain[, 1:(dim(Datain)[2]-1)]
4 pdata <- princomp(X, cor=T)
5 pc.data <- pdata$scores
6 pc.data1 <- -1*pc.data[,1]
7 pc.data2 <- -1*pc.data[,2]
8 X <- cbind(pc.data1, pc.data2)
9 cl <- kmeans(X, 10)
10 cl$cluster
11 plot(pc.data1, pc.data2, col=cl$cluster)
12 points(cl$centers, pch=16)
```

The script performs the following steps: reads a data file, extracts the response variable (YY) and predictor variables (XX), performs principal component analysis (PCA) on XX, extracts the principal component scores (pc.data), and then uses K-means clustering (kmeans) on the transformed data (X) to identify 10 clusters. The results are visualized using a scatter plot (plot) and the cluster centers are marked with pch=16 (points).

### 3. See the cluster output in R console using command Cluster

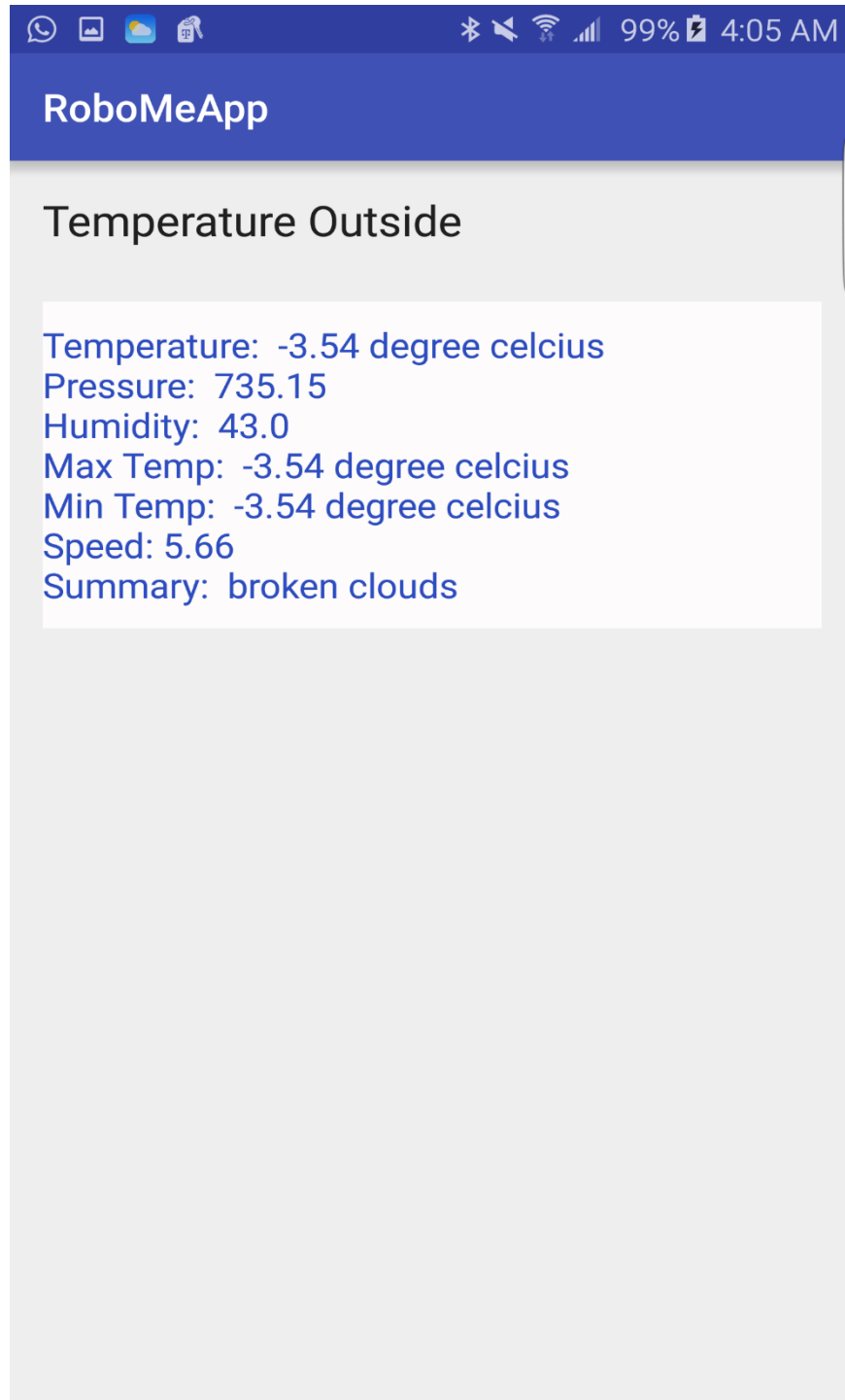
```
12:27 [Top Level] R Script
Console C:/Users/Rakesh/Desktop/Big Data Analytics and apps/Lab Assignments/lab2/
[681] 4 5 5 10 1 5 4 4 5 2 10 7 5 2 8 2 3 8 3 3 5 7 2 4 4 3 8 9 1 5 5 3 7 9
[715] 10 2 6 1 5 3 10 5 3 3 10 3 5 9 10 4 3 10 6 9 10 9 9 10 9 10 2 4 4 7 2 7 5 5
[749] 3 1 3 5 4 6 7 5 3 10 4 7 9 7 10 7 9 3 10 9
> plot(pc.comp1, pc.comp2,col=c1$cluster)
> points(c1$centers, pch=16)
> Datain <- read.table(file.choose(),sep = ",",header=FALSE)
> YY <- Datain[,dim(Datain)[2]]
> XX <- Datain[,1:(dim(Datain)[2]-1)]
> pdata <- princomp(X, cor=T)
> pc.data <- pdata$scores
> pc.data1 <- -1*pc.data[,1]
> pc.data2 <- -1*pc.data[,2]
> X <- cbind(pc.data1, pc.data2)
> c1 <- kmeans(X,10)
> c1$cluster
[1] 1 5 3 5 2 3 5 3 4 8 3 1 8 2 1 7 2 3 6 6 2 8 1 10 1 1 8 5 8 3 1 9 5 3
[35] 1 6 8 1 6 4 10 1 8 4 3 2 5 5 10 7 5 5 5 4 4 7 4 2 10 9 7 3 7 6 3 3 10 1
[69] 5 10 6 10 8 9 6 7 3 10 5 5 5 7 10 5 1 6 1 6 1 5 7 10 10 8 6 10 5 5 5 2 10 5
[103] 5 5 5 6 3 10 6 6 9 4 6 5 10 8 3 5 5 5 2 10 6 8 5 9 9 6 10 3 10 3 9 10 5 6
[137] 6 6 5 9 3 10 6 8 9 7 3 9 8 5 9 3 4 2 1 4 5 6 5 1 10 10 9 5 10 10 6 3 3 10
[171] 3 10 5 9 5 4 3 2 1 10 3 6 5 7 3 1 4 9 10 10 7 10 3 3 3 4 7 5 9 9 6 10 6 5
[205] 1 5 4 1 6 1 5 9 1 9 10 4 9 10 10 3 2 1 7 1 5 6 5 9 2 9 10 4 6 3 5 10 1 9
[239] 1 5 5 6 5 10 9 1 8 2 4 5 3 3 5 6 1 6 6 5 9 1 10 5 6 8 3 10 5 9 7 7 4 5
[273] 3 6 8 9 3 6 8 6 5 1 1 8 3 1 2 2 5 10 6 6 2 9 3 10 9 9 1 8 5 6 10 10 3 9
[307] 1 6 9 9 3 9 10 5 10 6 5 3 9 8 10 6 5 1 6 6 9 8 9 10 1 5 5 8 5 2 5 3 4 1
[341] 6 6 5 3 8 1 6 7 5 5 3 3 3 5 5 8 9 3 1 2 4 8 1 8 9 6 3 5 5 10 2 5 6 6
[375] 9 4 5 6 10 9 6 5 6 6 6 5 10 1 1 6 10 10 6 3 10 9 6 6 7 10 3 8 10 3 3 9 3 7
[409] 1 2 10 9 2 6 9 2 6 10 7 6 9 5 9 5 4 4 7 10 9 10 7 10 5 3 5 5 1 3 5 3 10 6
[443] 6 3 3 2 5 9 6 6 5 5 9 3 6 1 3 6 4 8 8 7 10 3 8 5 5 6 3 4 9 6 6 8 5 10
[477] 9 3 10 1 9 6 5 6 5 9 2 4 5 8 6 10 10 10 7 8 3 5 1 4 6 6 6 10 10 3 10 6 6 8
[511] 8 6 8 5 5 10 1 8 8 1 5 9 7 8 5 5 5 5 6 5 6 6 9 7 6 7 3 3 9 2 4 9 1 10
[545] 6 4 4 10 10 1 5 6 8 5 6 10 6 8 1 3 8 2 6 3 5 5 6 10 10 6 3 5 6 6 9 6 10 6
[579] 8 4 9 5 8 3 4 5 8 3 4 7 1 9 3 6 9 9 3 5 3 5 5 7 6 1 3 6 2 5 2 5 6 10
[613] 4 10 1 5 3 7 1 7 9 6 1 6 5 9 5 5 3 5 3 6 5 6 3 8 3 6 10 5 6 3 8 7 6 2
[647] 10 9 1 5 5 6 10 5 6 2 5 2 8 9 8 2 4 4 10 9 8 3 10 1 1 5 1 2 8 3 8 5 5 6
[681] 5 9 9 3 8 9 5 5 9 4 3 8 9 4 7 4 10 7 10 10 9 1 4 5 5 10 7 6 8 9 9 10 1 6
[715] 3 4 2 8 9 1 3 9 10 10 3 10 9 6 3 5 10 3 2 6 3 6 6 3 6 3 4 5 5 8 4 1 9 9
[749] 10 8 10 9 5 2 1 9 1 3 5 8 6 1 3 1 6 10 3 6
> plot(pc.data1, pc.data2,col=c1$cluster)
> points(c1$centers, pch=16)
>
```

4. Graph of cluster output with 10 clusters and Dark black dots indicate the centroid of means of cluster.



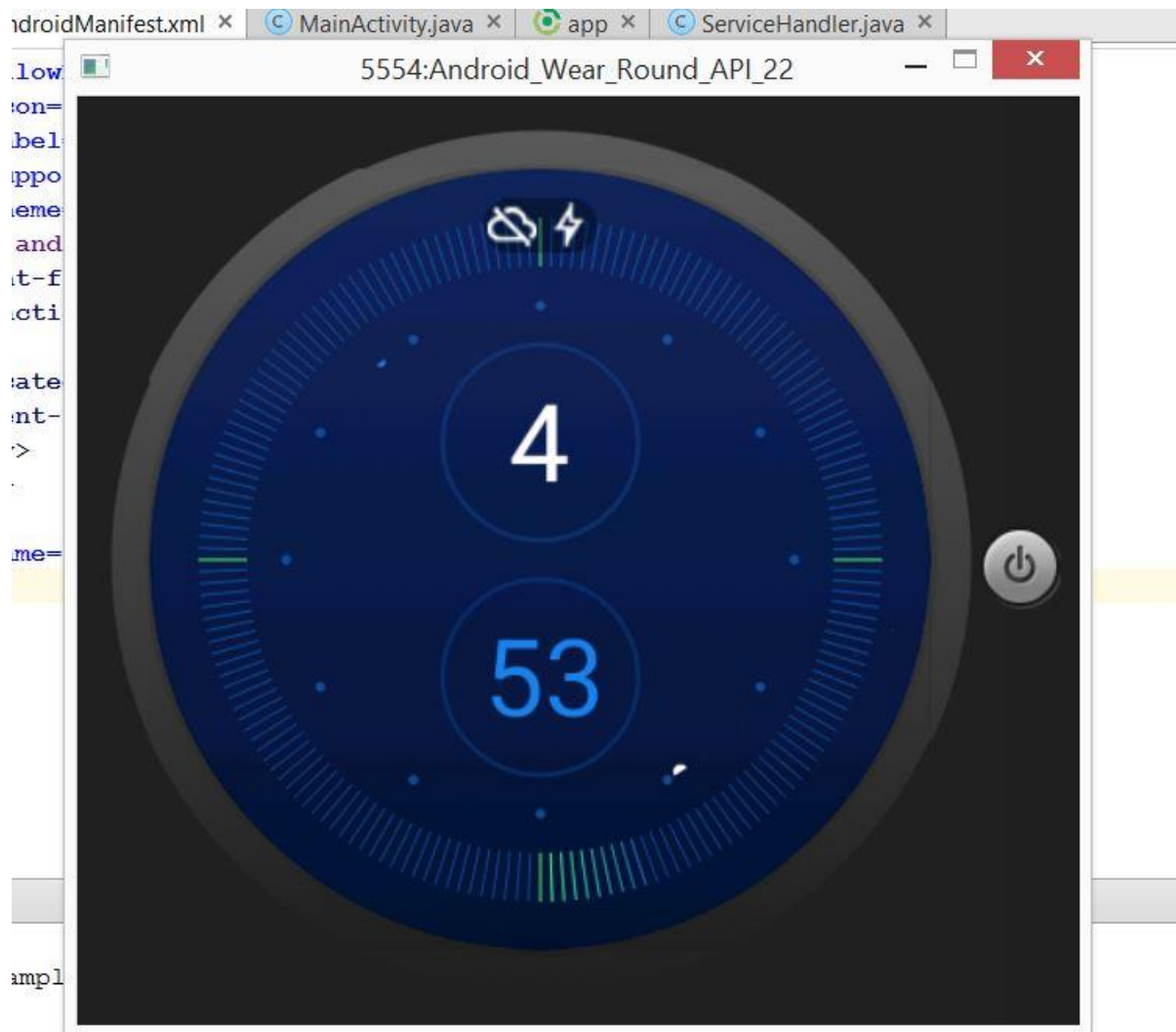
## 2. ROBOME AND SMART WATCH APP

In the Robome phone, the application output. This application sends temperature outside based on the latitude and longitude of the Robot to the Smart watch.



## Android wear side:

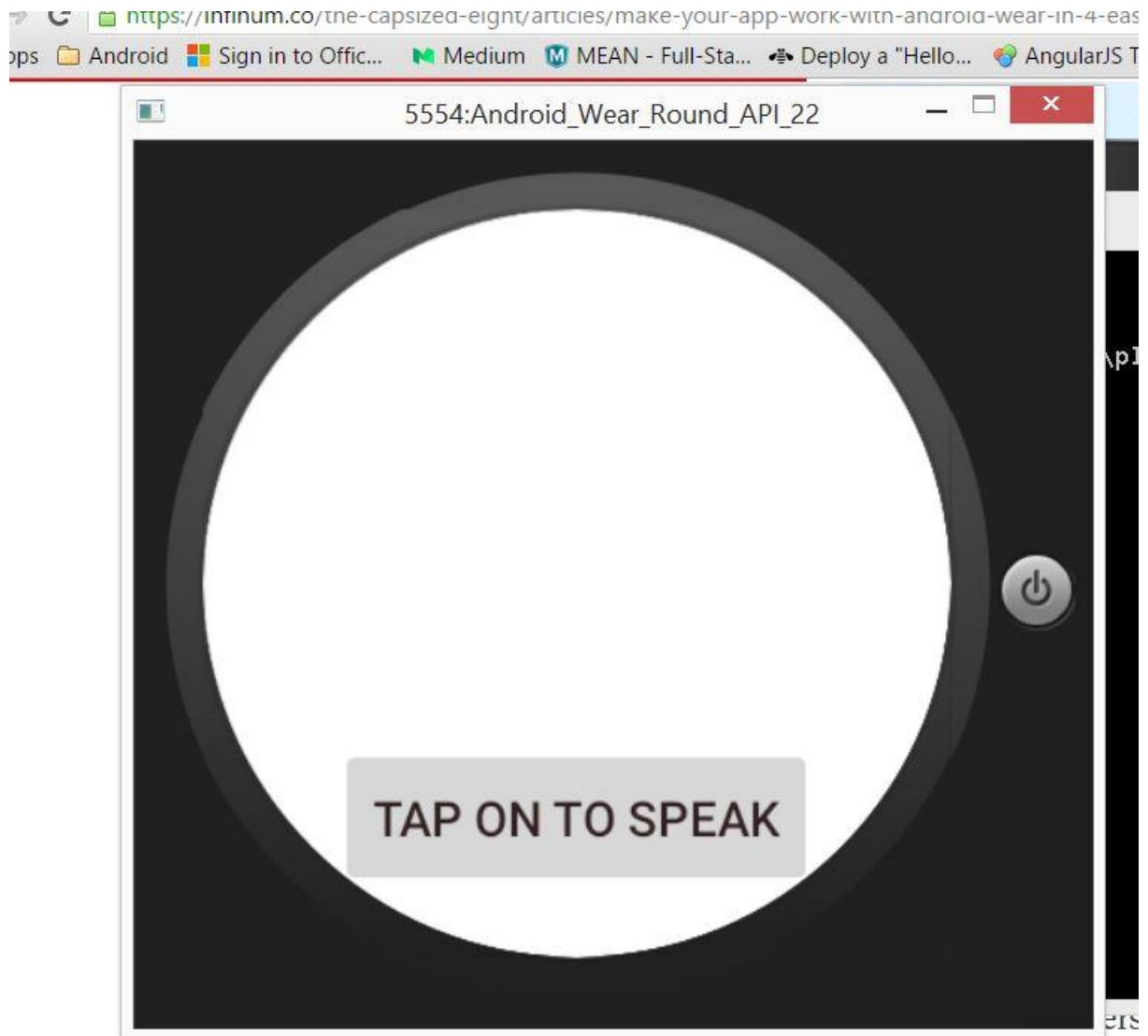
The android wear application is used to connect the phone and smart watch.



Can see the temperature on smart watch running on device.



This is android speech to text application in android wear which is used to send request to android phone.



wrists with informati

Notifications are a co



Here speak something example: Know the location temperature.

