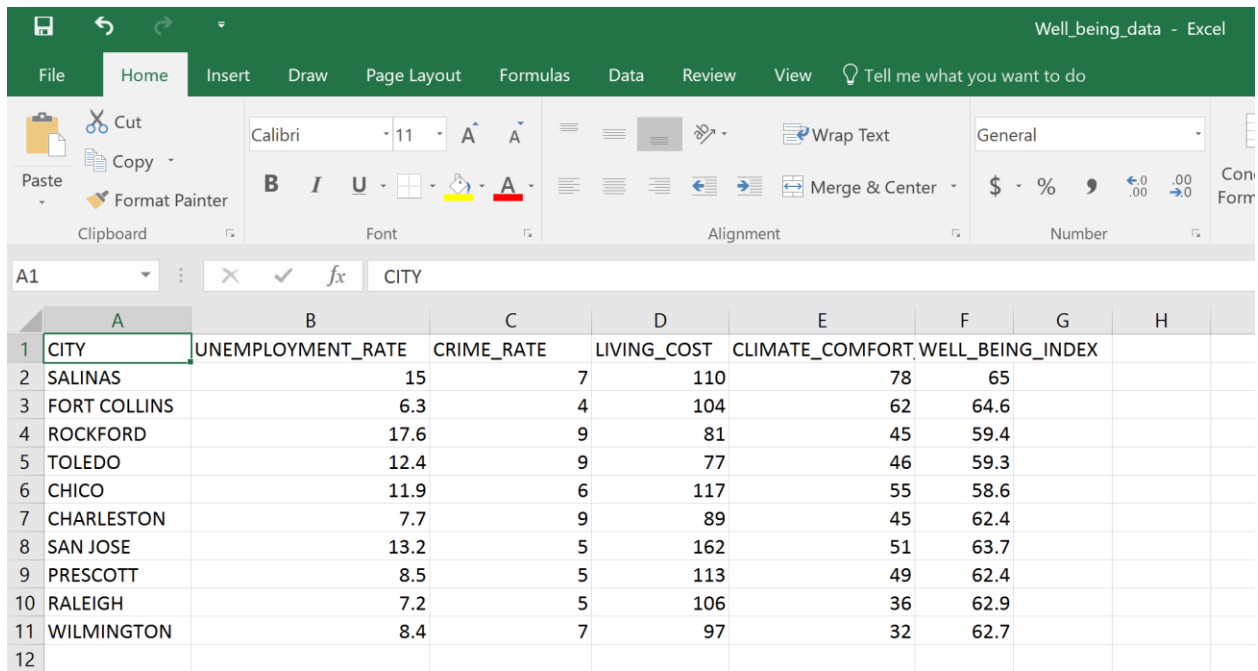


HOMEWORK 1

Data:



	A	B	C	D	E	F	G	H
1	CITY	UNEMPLOYMENT_RATE	CRIME_RATE	LIVING_COST	CLIMATE_COMFORT	WELL_BEING_INDEX		
2	SALINAS	15	7	110	78	65		
3	FORT COLLINS	6.3	4	104	62	64.6		
4	ROCKFORD	17.6	9	81	45	59.4		
5	TOLEDO	12.4	9	77	46	59.3		
6	CHICO	11.9	6	117	55	58.6		
7	CHARLESTON	7.7	9	89	45	62.4		
8	SAN JOSE	13.2	5	162	51	63.7		
9	PRESCOTT	8.5	5	113	49	62.4		
10	RALEIGH	7.2	5	106	36	62.9		
11	WILMINGTON	8.4	7	97	32	62.7		
12								

Code:

```
df="C:\\Users\\abinaya\\desktop\\Well_being_data.csv"
```

```
dataframe <- read.csv(df, header=TRUE, sep=",", as.is=TRUE)
```

```
plot(dataframe$UNEMPLOYMENT_RATE,dataframe$WELL_BEING_INDEX,main="Scatterplot  
of unemployment rate vs well being index",
```

```
      xlab="Unemployment rate", ylab="well being ")
```

```
plot(dataframe$CRIME_RATE,dataframe$WELL_BEING_INDEX, main="Scatterplot of crime rate vs well  
being index",
```

```
      xlab="crime rate", ylab="well being ")
```

```
plot(dataframe$LIVING_COST,dataframe$WELL_BEING_INDEX, main="Scatterplot of living cost vs well  
being index",
```

```
      xlab="living cost", ylab="well being ")
```

```
plot(dataframe$CLIMATE_COMFORT_LEVEL,dataframe$WELL_BEING_INDEX, main="Scatterplot of  
climate comfort level vs well being index",
```

```
xlab="climate comfort level", ylab="well being ")
```

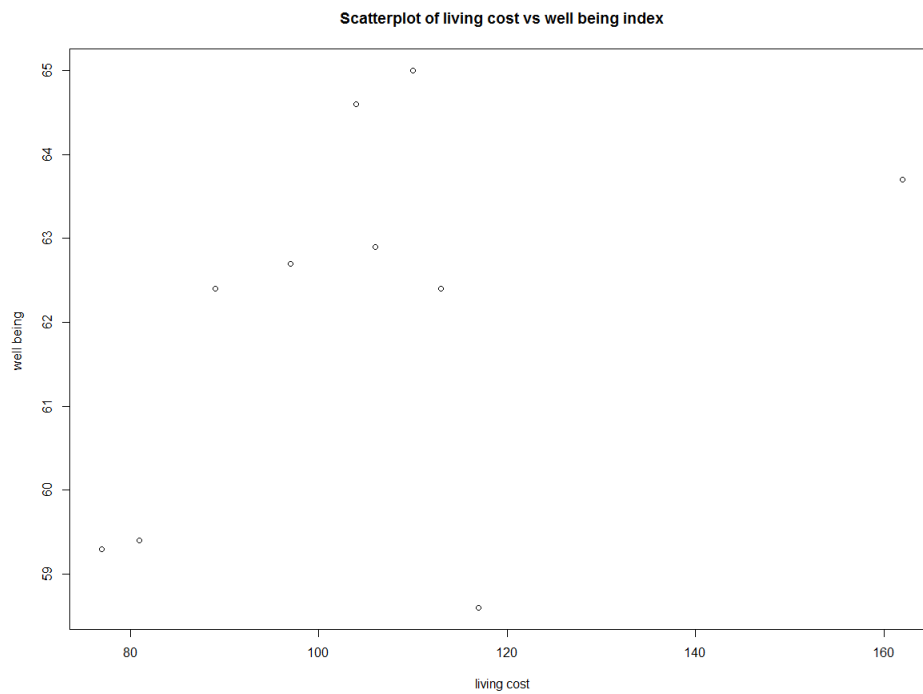
```
cor(dataframe$CLIMATE_COMFORT_LEVEL,dataframe$WELL_BEING_INDEX)
```

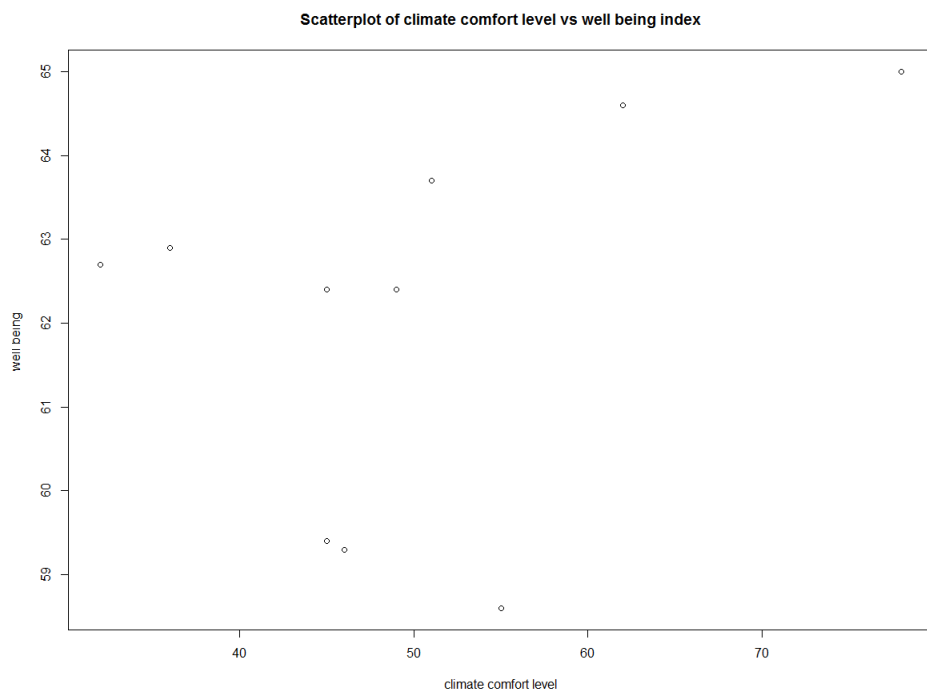
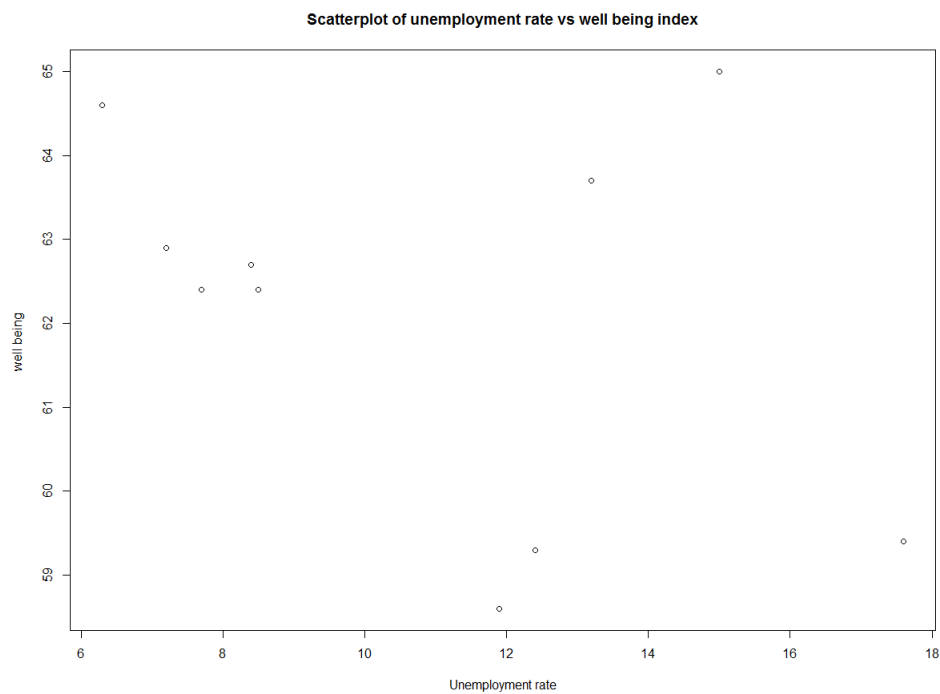
```
cor(dataframe$UNEMPLOYMENT_RATE,dataframe$WELL_BEING_INDEX)
```

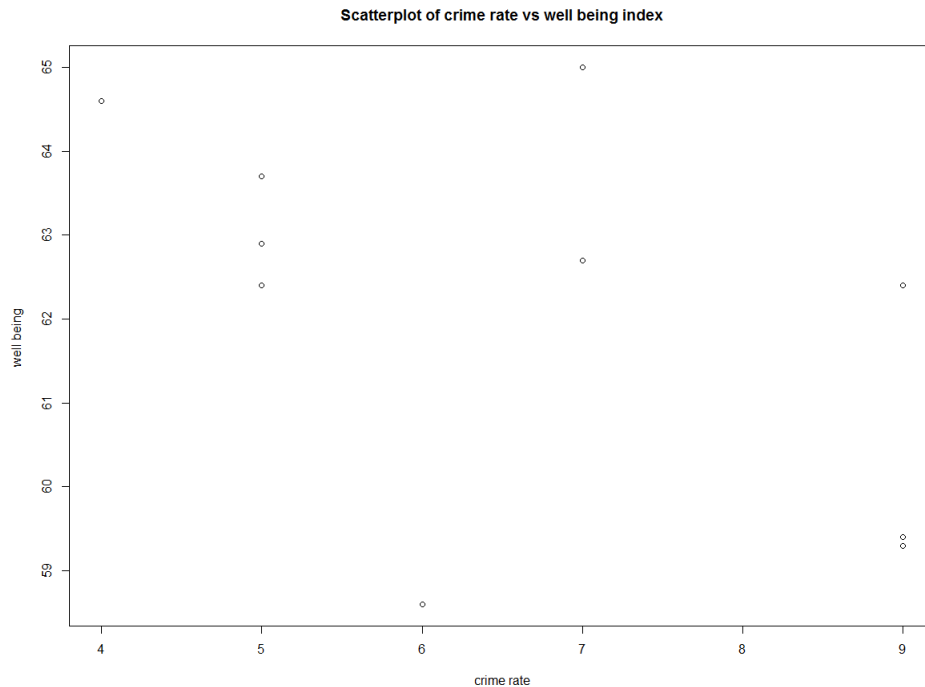
```
cor(dataframe$CRIME_RATE,dataframe$WELL_BEING_INDEX)
```

```
cor(dataframe$LIVING_COST,dataframe$WELL_BEING_INDEX)
```

PLOTS:







Analysis:

Correlation of living cost and well being index= 0.4073668

Correlation of unemployment rate and well being index= -0.3649807

Correlation of crime rate and well being index= -0.5141085

Correlation of climate comfort level and well being index= 0.3622013

From the correlation values we can find that the crime rate and unemployment rate have a negative correlation with the well being index. And the living cost and climate comfort level have a positive correlation with the well being index.

Since there are outliers these correlated values cannot help in making the absolute decision, So, let's make a paired comparison table to find out the weightage of each of these values. Let us rank each of the attributes based on its correlation value and calculate its percentage of importance

Paired comparison table:

	Living cost:L	Unemployment rate :U	Crime rate:C	Climate comfort level:CC
Living cost: L	NA	L,1	C,1	L,1
Unemployment rate :U	NA	NA	C,2	U,1
Crime rate : C	NA	NA	NA	C,2
Climate comfort level :CC	NA	NA	NA	

L: 2 (25 percent)
U:1 (12 percent)
C:5 (60 percent)
CC:0

So, crime rate constitutes of 60 percent of the well being index and the living cost constitutes of 25 percent of the prediction

Analysis:

Let us take the city Lubbock and predict its well being index based on its other attributes.

City	Unemployment rate	Crime rate	Living cost	Climate_comfort_level
LUBBOCK	5.3	8	76	35

Let us split the positive and negative correlated attributes

Positive	Negative
Living cost=76	Unemployment rate= 5.3
Climate comfort level =35	Crime rate =8

The positive attributes will try to increase the well being index while the negative index will try to decrease the value.

Since the crime rate contributes to 60 percentage of the prediction,Lets first compare the crime rate of Lubbock with the crime rate of other cities and find the closest match.The cities Rockford,Toledo and Charleston have a crime rate of 9.

Lets compare the living cost and crime rate with that of the other cities.The living cost and crime rate of other cities which are closest to Lubbock are Rockford,Toledo and Charleston.

Now lets compare Unemployment rate along with the other two features. The cities Rockford and Toledo have a high unemployment rate than Lubbock so lets eliminate those cities.The city Charleston have a closer value of unemployment like Lubbock .

So, Charleston have a similar attributes like Lubbock and It has a well being index of 62.4. But it has a higher climate comfort level than Lubbock so the well being of Lubbock will be lower than 62.4.So, the well being index of Lubbock will be in the range of **58-60**

Steps to improve:

We can improve our prediction by the following ways.

- By assessing a large number of attributes which may directly or indirectly affect our prediction

- By eliminating the inaccuracy in data
- By using a better predictive model
- Normalizing values.(cost of living may change from place to place)
- By analyzing the outliers.

Discussion:

The prediction method used above is not accurate and the well being index may vary from the predicted range. Since each of the attribute value is dependent on the other attributes as well. Example, The unemployment rate may depend upon the education level of the people, presence of opportunities and the several other factors. Estimating the well being is a holistic approach and we need to consider hundreds of factors before estimating the value.

Inaccuracy in data:

Inaccurate data would lead to a false prediction of value. Take for example not all crimes in a city is reported to the police station. Since we take crime rate as a major factor for predicting the well being index. If the given data is inaccurate it will cost a lot while making prediction. And the things like cost of living, transportation may vary from place to place. (cost of living in CA > IL). Eliminating the inaccurate data and using a large number of dataset attributes may cost a lot of time. By normalizing data and segregating the redundant information the prediction can be improved.

