NAME:Aditya Somani CLASS:TE-A ROLL NO:T1851061

ASSIGNMENT-2

Part B: Assignments based on R and Python Aim: Perform the following operations using R/Python on the Air quality and Heart Diseases data sets

1) Data cleaning

2) Data integration

3) Data transformation

4) Error correcting

5) Data model building

Introduction.

The Air Quality Dataset is available in R Studio and the Operation are performed on this data.

Aim of analysis

In the following document, 4 different machine learning algorithms to predict heart disease (angiographic disease status) are compared. For some algorithms, model parameters are tuned and the best model selected. The best results measured by AUC and accuracy are obtained from a logistic regression model (AUC 0.92, Accuracy 0.87), followed by Gradient Boosting Machines. From a set of 14 variables, the most important to predict heart failure are whether or not there is a reversable defect in Thalassemia followed by whether or not there is an occurrence of asymptomatic chest pain.

Dataset:

Nicely prepared heart disease data are available at UCI. The document mentions that previous work resulted in an accuracy of 74-77% for the preciction of heart disease using the cleveland data.

Variable name	**Short desciption**	**Variable name**	**Short description**
age	Age of patient	thalach	maximum heart rate achieved
sex	Sex, 1 for male	exang	exercise induced angina (1 yes)
cp	chest pain	oldpeak	ST depression induc. ex.
trestbps	resting blood pressure	slope	slope of peak exercise ST
chol	serum cholesterol	ca	number of major vessel
fbs	fasting blood sugar larger 120mg/dl (1 true)	thal	no explanation provided, but probably thalassemia (3 normal; 6 fixed defect; 7 reversable defect)

restecg	resting electroc. result (1 anomality)	num	diagnosis of heart disease (angiographic disease status)

The variable we want to predict is **num** with Value 0: < 50% diameter narrowing and Value 1: > 50% diameter narrowing. We assume that every value with 0 means heart is okay, and 1,2,3,4 means heart disease.

From the possible values the variables can take, it is evident that the following need to be dummified because the distances in the values is random: cp,thal, restecg, slope.

Operations Performed on Air Quality Dataset

Read the Dataset

```
> airquality
   Ozone Solar.R Wind Temp Month Day
1     41     190  7.4   67     5   1
2     36     118  8.0   72     5   2
3     12     149 12.6   74     5   3
4     18     313 11.5   62     5   4
5     NA      NA 14.3   56     5   5
6     28      NA 14.9   66     5   6
7     23     299  8.6   65     5   7
8     19      99 13.8   59     5   8
9      8      19 20.1   61     5   9
10    NA     194  8.6   69     5  10
11     7      NA  6.9   74     5  11
12    16     256  9.7   69     5  12
13    11     290  9.2   66     5  13
14    14     274 10.9   68     5  14
15    18      65 13.2   58     5  15
```

View the Summary of Data

```
> summary(airquality)
     Ozone           Solar.R           Wind             Temp
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00
 NA's   :37       NA's   :7
     Month            Day
 Min.   :5.000   Min.   : 1.0
 1st Qu.:6.000   1st Qu.: 8.0
 Median :7.000   Median :16.0
 Mean   :6.993   Mean   :15.8
 3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :9.000   Max.   :31.0
```

1) Data cleaning (Students have to Insert the screenshot/output of every options operations Performed by them.)

Conclusion: Thus we have learnt various operations of 1) Data cleaning 2) Data integration 3) Data transformation 4) Error correcting 5) Data model building) with **R Language in RStudio**.

DATA CLEANING

```
> summary(airquality)

    Ozone        Solar.R        Wind          Temp         Month

 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000

 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000

 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00   Median :7.000

 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88   Mean   :6.993

 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000

 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000

 NA's   :37       NA's   :7

     Day

 Min.   : 1.0

 1st Qu.: 8.0

 Median :16.0

 Mean   :15.8

 3rd Qu.:23.0
```

```
 Max.   :31.0

> air=airquality
> air$Ozone=ifelse(is.na(air$Ozone),median(air$Ozone,na.rm = TRUE),air$Ozone)
> summary(air)
    Ozone          Solar.R          Wind            Temp           Month
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000
 1st Qu.: 21.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00   Median :7.000
 Mean   : 39.56   Mean   :185.9   Mean   : 9.958   Mean   :77.88   Mean   :6.993
 3rd Qu.: 46.00   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000
                  NA's   :7
      Day
 Min.   : 1.0
 1st Qu.: 8.0
 Median :16.0
 Mean   :15.8
 3rd Qu.:23.0
 Max.   :31.0

> air$Solar.R=ifelse(is.na(air$Solar.R),median(air$Solar.R,na.rm = TRUE),air$Solar.R)
> summary(air)
    Ozone          Solar.R          Wind            Temp           Month
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000
 1st Qu.: 21.00   1st Qu.:120.0   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00   Median :7.000
```

```
Mean   : 39.56   Mean   :186.8   Mean   : 9.958   Mean   :77.88   Mean   :6.993
3rd Qu.: 46.00   3rd Qu.:256.0   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000
Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000
     Day
Min.   : 1.0
1st Qu.: 8.0
Median :16.0
Mean   :15.8
3rd Qu.:23.0
Max.   :31.0
```

DATA TRANSFORMATION

```
> brks=c(0,50,100,150,200,250,300,350)
> air$Solar.R=cut(air$Solar.R,breaks = brks,include.lowest = TRUE)
> head(air)
   Ozone   Solar.R Wind Temp Month Day
1    41.0 (150,200]  7.4   67     5   1
2    36.0 (100,150]  8.0   72     5   2
3    12.0 (100,150] 12.6   74     5   3
4    18.0 (300,350] 11.5   62     5   4
5    31.5 (200,250] 14.3   56     5   5
6    28.0 (200,250] 14.9   66     5   6
> tem=c(0,15,30,45,60,75,90,105)
> air$Temp=cut(air$Temp,breaks = tem,include.lowest = TRUE)
> head(air)
  Ozone   Solar.R Wind    Temp Month Day
1    41.0 (100,200]  7.4 (60,75]     5   1
```

```
2    36.0 (100,200]  8.0 (60,75]     5   2
3    12.0 (100,200] 12.6 (60,75]     5   3
4    18.0 (300,400] 11.5 (60,75]     5   4
5    31.5 (200,300] 14.3 (45,60]     5   5
6    28.0 (200,300] 14.9 (60,75]     5   6
> air$Month=gsub(5,"May",air$Month)
> head(air)
  Ozone   Solar.R Wind    Temp Month Day
1    41.0 (100,200]  7.4 (60,75]   May   1
2    36.0 (100,200]  8.0 (60,75]   May   2
3    12.0 (100,200] 12.6 (60,75]   May   3
4    18.0 (300,400] 11.5 (60,75]   May   4
5    31.5 (200,300] 14.3 (45,60]   May   5
6    28.0 (200,300] 14.9 (60,75]   May   6
> air$Month=gsub(6,"June",air$Month)
> air$Month=gsub(7,"July",air$Month)
> air$Month=gsub(8,"August",air$Month)
> air$Month=gsub(9,"Sept",air$Month)
> head(air)
  Ozone   Solar.R Wind    Temp Month Day
1    41.0 (100,200]  7.4 (60,75]   May   1
2    36.0 (100,200]  8.0 (60,75]   May   2
3    12.0 (100,200] 12.6 (60,75]   May   3
4    18.0 (300,400] 11.5 (60,75]   May   4
5    31.5 (200,300] 14.3 (45,60]   May   5
```

```
6    28.0 (200,300] 14.9 (60,75]   May   6
> air
   Ozone   Solar.R Wind     Temp  Month Day
1        41.0 (100,200]  7.4  (60,75]    May   1
2        36.0 (100,200]  8.0  (60,75]    May   2
3        12.0 (100,200] 12.6  (60,75]    May   3
4        18.0 (300,400] 11.5  (60,75]    May   4
5        31.5 (200,300] 14.3  (45,60]    May   5
6        28.0 (200,300] 14.9  (60,75]    May   6
7        23.0 (200,300]  8.6  (60,75]    May   7
8        19.0   [0,100] 13.8  (45,60]    May   8
9        8.0   [0,100] 20.1  (60,75]    May   9 10   31.5 (100,200]  8.6  (60,75]    May  10
11        7.0 (200,300]  6.9  (60,75]    May  11
12        16.0 (200,300]  9.7  (60,75]    May  12
13        11.0 (200,300]  9.2  (60,75]    May  13
14        14.0 (200,300] 10.9  (60,75]    May  14
15        18.0   [0,100] 13.2  (45,60]    May  15
16        14.0 (300,400] 11.5  (60,75]    May  16
17        34.0 (300,400] 12.0  (60,75]    May  17
18        6.0   [0,100] 18.4  (45,60]    May  18
19        30.0 (300,400] 11.5  (60,75]    May  19
20        11.0   [0,100]  9.7  (60,75]    May  20
21        1.0   [0,100]  9.7  (45,60]    May  21
22        11.0 (300,400] 16.6  (60,75]    May  22
23        4.0   [0,100]  9.7  (60,75]    May  23
```

24 32.0 [0,100] 12.0 (60,75] May 24

25 31.5 [0,100] 16.6 (45,60] May 25

26 31.5 (200,300] 14.9 (45,60] May 26

27 31.5 (200,300] 8.0 (45,60] May 27

28 23.0 [0,100] 12.0 (60,75] May 28

29 45.0 (200,300] 14.9 (75,90] May 29

30 115.0 (200,300] 5.7 (75,90] May 30

31 37.0 (200,300] 7.4 (75,90] May 31

32 31.5 (200,300] 8.6 (75,90] June 1

33 31.5 (200,300] 9.7 (60,75] June 2

34 31.5 (200,300] 16.1 (60,75] June 3

35 31.5 (100,200] 9.2 (75,90] June 4

36 31.5 (200,300] 8.6 (75,90] June 5

37 31.5 (200,300] 14.3 (75,90] June 6

38 29.0 (100,200] 9.7 (75,90] June 7

39 31.5 (200,300] 6.9 (75,90] June 8

40 71.0 (200,300] 13.8 (75,90] June 9 41 39.0 (300,400] 11.5 (75,90] June 10

42 31.5 (200,300] 10.9 (90,105] June 11

43 31.5 (200,300] 9.2 (90,105] June 12

44 23.0 (100,200] 8.0 (75,90] June 13

45 31.5 (300,400] 13.8 (75,90] June 14

46 31.5 (300,400] 11.5 (75,90] June 15

47 21.0 (100,200] 14.9 (75,90] June 16

48 37.0 (200,300] 20.7 (60,75] June 17

49 20.0 [0,100] 9.2 (60,75] June 18

50 12.0 (100,200] 11.5 (60,75] June 19

51 13.0 (100,200] 10.3 (75,90] June 20

52 31.5 (100,200] 6.3 (75,90] June 21

53 31.5 [0,100] 1.7 (75,90] June 22

54 31.5 [0,100] 4.6 (75,90] June 23

55 31.5 (200,300] 6.3 (75,90] June 24

56 31.5 (100,200] 8.0 (60,75] June 25

57 31.5 (100,200] 8.0 (75,90] June 26

58 31.5 [0,100] 10.3 (60,75] June 27

59 31.5 [0,100] 11.5 (75,90] June 28

60 31.5 [0,100] 14.9 (75,90] June 29

61 31.5 (100,200] 8.0 (75,90] June 30

62 135.0 (200,300] 4.1 (75,90] July 1

63 49.0 (200,300] 9.2 (75,90] July 2

64 32.0 (200,300] 9.2 (75,90] July 3

65 31.5 (100,200] 10.9 (75,90] July 4

66 64.0 (100,200] 4.6 (75,90] July 5

67 40.0 (300,400] 10.9 (75,90] July 6

68 77.0 (200,300] 5.1 (75,90] July 7

69 97.0 (200,300] 6.3 (90,105] July 8

70 97.0 (200,300] 5.7 (90,105] July 9

71 85.0 (100,200] 7.4 (75,90] July 10

72 31.5 (100,200] 8.6 (75,90] July 11 73 10.0 (200,300] 14.3 (60,75] July 12

74 27.0 (100,200] 14.9 (75,90] July 13

75 31.5 (200,300] 14.9 (90,105] July 14

76 7.0 [0,100] 14.3 (75,90] July 15

77 48.0 (200,300] 6.9 (75,90] July 16

78 35.0 (200,300] 10.3 (75,90] July 17

79 61.0 (200,300] 6.3 (75,90] July 18

80 79.0 (100,200] 5.1 (75,90] July 19

81 63.0 (200,300] 11.5 (75,90] July 20

82 16.0 [0,100] 6.9 (60,75] July 21

83 31.5 (200,300] 9.7 (75,90] July 22

84 31.5 (200,300] 11.5 (75,90] July 23

85 80.0 (200,300] 8.6 (75,90] July 24

86 108.0 (200,300] 8.0 (75,90] July 25

87 20.0 [0,100] 8.6 (75,90] July 26

88 52.0 [0,100] 12.0 (75,90] July 27

89 82.0 (200,300] 7.4 (75,90] July 28

90 50.0 (200,300] 7.4 (75,90] July 29

91 64.0 (200,300] 7.4 (75,90] July 30

92 59.0 (200,300] 9.2 (75,90] July 31

93 39.0 [0,100] 6.9 (75,90] August 1

94 9.0 [0,100] 13.8 (75,90] August 2

95 16.0 [0,100] 7.4 (75,90] August 3

96 78.0 (200,300] 6.9 (75,90] August 4

97 35.0 (200,300] 7.4 (75,90] August 5 98 66.0 (200,300] 4.6 (75,90] August 6

99 122.0 (200,300] 4.0 (75,90] August 7

100 89.0 (200,300] 10.3 (75,90] August 8

101 110.0 (200,300] 8.0 (75,90] August 9

102 31.5 (200,300] 8.6 (90,105] August 10

103 31.5 (100,200] 11.5 (75,90] August 11

104 44.0 (100,200] 11.5 (75,90] August 12

105 28.0 (200,300] 11.5 (75,90] August 13

106 65.0 (100,200] 9.7 (75,90] August 14

107 31.5 [0,100] 11.5 (75,90] August 15

108 22.0 [0,100] 10.3 (75,90] August 16

109 59.0 [0,100] 6.3 (75,90] August 17

110 23.0 (100,200] 7.4 (75,90] August 18

111 31.0 (200,300] 10.9 (75,90] August 19

112 44.0 (100,200] 10.3 (75,90] August 20

113 21.0 (200,300] 15.5 (75,90] August 21

114 9.0 [0,100] 14.3 (60,75] August 22

115 31.5 (200,300] 12.6 (60,75] August 23

116 45.0 (200,300] 9.7 (75,90] August 24

117 168.0 (200,300] 3.4 (75,90] August 25

118 73.0 (200,300] 8.0 (75,90] August 26

119 31.5 (100,200] 5.7 (75,90] August 27

120 76.0 (200,300] 9.7 (90,105] August 28

121 118.0 (200,300] 2.3 (90,105] August 29

122 84.0 (200,300] 6.3 (90,105] August 30 123 85.0 (100,200] 6.3 (90,105] August 31

124 96.0 (100,200] 6.9 (90,105] Sept 1

125 78.0 (100,200] 5.1 (90,105] Sept 2

126 73.0 (100,200] 2.8 (90,105] Sept 3

127 91.0 (100,200] 4.6 (90,105] Sept 4

128 47.0 [0,100] 7.4 (75,90] Sept 5

129 32.0 [0,100] 15.5 (75,90] Sept 6

130 20.0 (200,300] 10.9 (75,90] Sept 7

131 23.0 (200,300] 10.3 (75,90] Sept 8

132 21.0 (200,300] 10.9 (60,75] Sept 9

133 24.0 (200,300] 9.7 (60,75] Sept 10

134 44.0 (200,300] 14.9 (75,90] Sept 11

135 21.0 (200,300] 15.5 (75,90] Sept 12

136 28.0 (200,300] 6.3 (75,90] Sept 13

137 9.0 [0,100] 10.9 (60,75] Sept 14

138 13.0 (100,200] 11.5 (60,75] Sept 15

139 46.0 (200,300] 6.9 (75,90] Sept 16

140 18.0 (200,300] 13.8 (60,75] Sept 17

141 13.0 [0,100] 10.3 (75,90] Sept 18

142 24.0 (200,300] 10.3 (60,75] Sept 19

143 16.0 (200,300] 8.0 (75,90] Sept 20

144 13.0 (200,300] 12.6 (60,75] Sept 21

145 23.0 [0,100] 9.2 (60,75] Sept 22

146 36.0 (100,200] 10.3 (75,90] Sept 23

147 7.0 [0,100] 10.3 (60,75] Sept 24 148 14.0 [0,100] 16.6 (60,75] Sept 25

149 30.0 (100,200] 6.9 (60,75] Sept 26

150 31.5 (100,200] 13.2 (75,90] Sept 27

151 14.0 (100,200] 14.3 (60,75] Sept 28

152 18.0 (100,200] 8.0 (75,90] Sept 29

153 20.0 (200,300] 11.5 (60,75] Sept 30

DATA MODEL

```
> for(i in 1:nrow(air1)){
+    if(is.na(air1[i,"Ozone"])){
+        air1[i,"Ozone"]<- mean(air1[which(air1[,"Month"]==air1[i,"Month"]),"Ozone"],na.rm
=TRUE)
+    }
+    #input monthly mean in Solar.R
+    if(is.na(air1[i,"Solar.R"])){
+        air1[i,"Solar.R"]<- mean(air1[which(air1[,"Month"]==air1[i,"Month"]),"Solar.R"],na.rm =
TRUE)
+    }
+ }
> summary(air1)
     Ozone          Solar.R          Wind            Temp           Month            Day
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00   Min.   :5.000   Min.
: 1.0
 1st Qu.: 21.00   1st Qu.:120.0   1st Qu.: 7.400   1st Qu.:72.00   1st Qu.:6.000   1st Qu.: 8.0
 Median : 29.44   Median :194.0   Median : 9.700   Median :79.00   Median :7.000   Median
:16.0
 Mean   : 40.85   Mean   :185.5   Mean   : 9.958   Mean   :77.88   Mean   :6.993   Mean   :15.8
 3rd Qu.: 59.12   3rd Qu.:256.0   3rd Qu.:11.500   3rd Qu.:85.00   3rd Qu.:8.000   3rd Qu.:23.0
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00   Max.   :9.000   Max.
:31.0
> head(air1)
     Ozone  Solar.R Wind Temp Month Day
```

1 41.00000 190.0000 7.4 67 5 1

2 36.00000 118.0000 8.0 72 5 2

3 12.00000 149.0000 12.6 74 5 3

4 18.00000 313.0000 11.5 62 5 4

5 23.61538 181.2963 14.3 56 5 5 6 28.00000 181.2963 14.9 66 5 6

> normalize<-function(x){

+ return(x-min(x)/(max(x)-min(x)))}

> air1<-normalize(air1)

> summary(air1)

Ozone Solar.R Wind Temp Month Day

Min. : 0.997 Min. : 6.997 Min. : 1.697 Min. :56.00 Min. :4.997 Min.
: 0.997

1st Qu.: 20.997 1st Qu.:119.997 1st Qu.: 7.397 1st Qu.:72.00 1st Qu.:5.997 1st Qu.:
7.997

Median : 29.441 Median :193.997 Median : 9.697 Median :79.00 Median :6.997 Median :15.997

Mean : 40.848 Mean :185.531 Mean : 9.955 Mean :77.88 Mean :6.990
Mean :15.801

3rd Qu.: 59.112 3rd Qu.:255.997 3rd Qu.:11.497 3rd Qu.:85.00 3rd Qu.:7.997 3rd Qu.:22.997

Max. :167.997 Max. :333.997 Max. :20.697 Max. :97.00 Max. :8.997 Max. :30.997

> normalize<-function(x){

+ return(x-min(x)/(max(x)-min(x)))

+ }

> air<-normalize(air)

> summary(air)

Ozone Solar.R Wind Temp Month

Min. : NA Min. : NA Min. : NA Min. : NA Min. : NA

1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA 1st Qu.: NA

Median : NA Median : NA Median : NA Median : NA Median : NA

Mean :NaN Mean :NaN Mean :NaN Mean :NaN Mean :NaN

3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA 3rd Qu.: NA

Max. : NA Max. : NA Max. : NA Max. : NA Max. : NA

NA's :153 NA's :153 NA's :153 NA's :153 NA's :153

Day

Min. : NA

1st Qu.: NA

Median : NA

Mean :NaN

3rd Qu.: NA

Max. : NA

NA's :153

> str(air)

'data.frame': 153 obs. of 6 variables:

$ Ozone : num NA NA NA NA NA NA NA NA NA NA ...

$ Solar.R: num NA NA NA NA NA NA NA NA NA NA ...

$ Wind : num NA NA NA NA NA NA NA NA NA NA ...

$ Temp : num NA NA NA NA NA NA NA NA NA NA ...

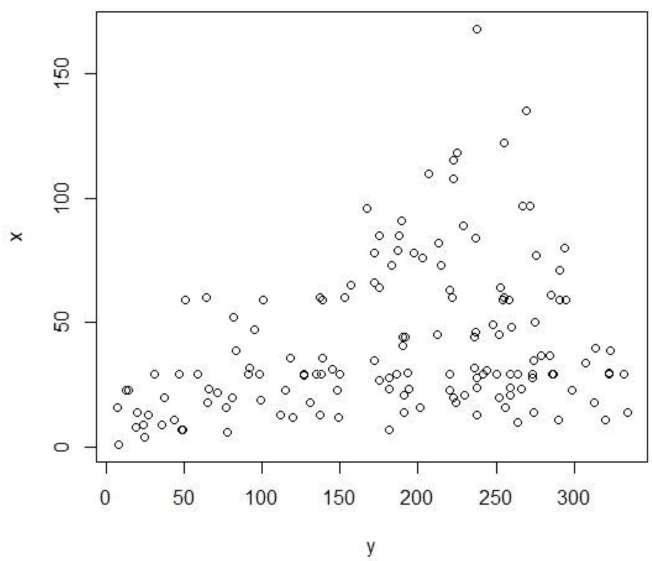
$ Month : num NA NA NA NA NA NA NA NA NA NA ...

$ Day : num NA NA NA NA NA NA NA NA NA NA ...

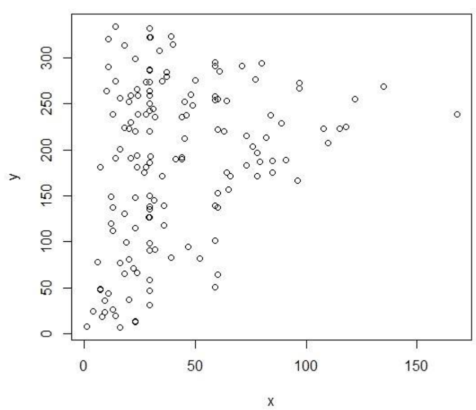
> x=air[,"Ozone"]

> y=air[,"Solar.R"]

> plot(x~y)



> plot(y~x)

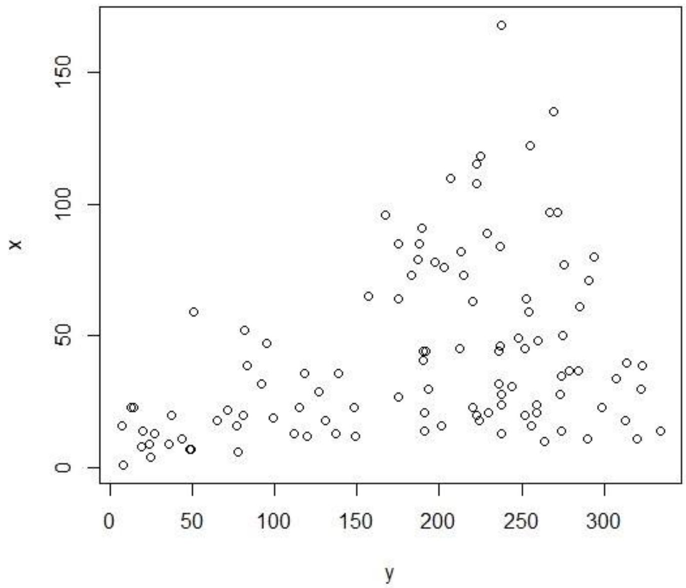


> model=lm(y~x)

> x=airquality[,"Ozone"]

> y=airquality[,"Solar.R"]

> plot(x~y)



> model1=lm(y~x)

> model1

Call: lm(formula =

y ~ x) Coefficients:

(Intercept) x

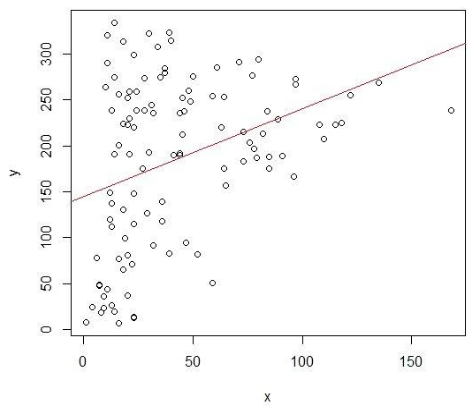
144.6306 0.9542

>

> abline(model1,col="pink")

> abline(model1,col="Blue")

> plot(y~x)



> model1=lm(y~x)

> model1

Call: lm(formula =

y ~ x) Coefficients:

(Intercept) x

144.6306 0.9542

> abline(model1,col="Red")

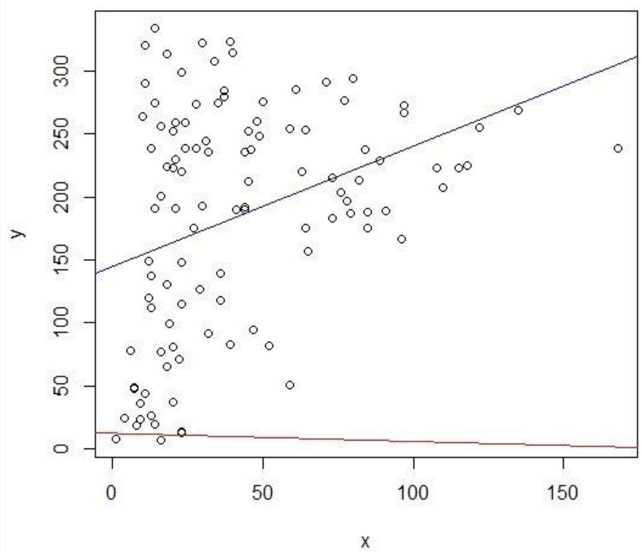
> abline(model1,co="Blue")

> x=airquality[,"Ozone"]

> y=airquality[,"Wind"]

> model1=lm(y~x)

> abline(model1,col="red")



DATA INTEGRATION

> Name=c("Kim Namjoon","Kim Seokjin","Min Yoongi","Jeon Hoseok","Park Jimin","Kim Taehyung","Jeon Jungkook")

> stage_name=c("RM","Jin","Suga","J-Hope","Jimin","V","JK")

> Songs=c("Moonchild","Epiphany","Seesaw","Chicken noodle soup","Filter","InnerChild","Euphoria")

> Bangtan=cbind(Name,Songs,stage_name)

> Bangtan

Name Songs stage_name

[1,] "Kim Namjoon" "Moonchild" "RM"

[2,] "Kim Seokjin" "Epiphany" "Jin"

[3,] "Min Yoongi" "Seesaw" "Suga"

```
[4,] "Jeon Hoseok"   "Chicken noodle soup" "J-Hope"

[5,] "Park Jimin"    "Filter"            "Jimin"

[6,] "Kim Taehyung"  "InnerChild"        "V"

[7,] "Jeon Jungkook" "Euphoria"          "JK"

> collabs=data.frame(Name=c("Ashley Nicolette Frangipane","Ari Staprans Leff","Maxwell
George Schneider,"),Songs=c("Boy with Luv","Make it
Right","Blueberry"),stage_name=c("Halsey","Lauv","MAX"),stringsAsFactors = FALSE)

> collabs

                   Name              Songs          stage_name

1 Ashley Nicolette Frangipane    Boy with Luv       Halsey

2         Ari Staprans Leff          Make it Right       Lauv

3   Maxwell George Schneider      Blueberry          MAX

> BTS=rbind(Bangtan,collabs)

> BTS

                   Name           Songs                      stage_name

1               Kim Namjoon         Moonchild                      RM

2               Kim Seokjin          Epiphany                       Jin

3                Min Yoongi          Seesaw                      Suga

4               Jeon Hoseok       Chicken noodle soup           J-Hope

5                Park Jimin              Filter                        Jimin

6              Kim Taehyung           InnerChild                       V

7             Jeon Jungkook           Euphoria                      JK

8  Ashley Nicolette Frangipane       Boy with Luv                Halsey

9          Ari Staprans Leff              Make it Right               Lauv

10   Maxwell George Schneider,        Blueberry              MAX
```