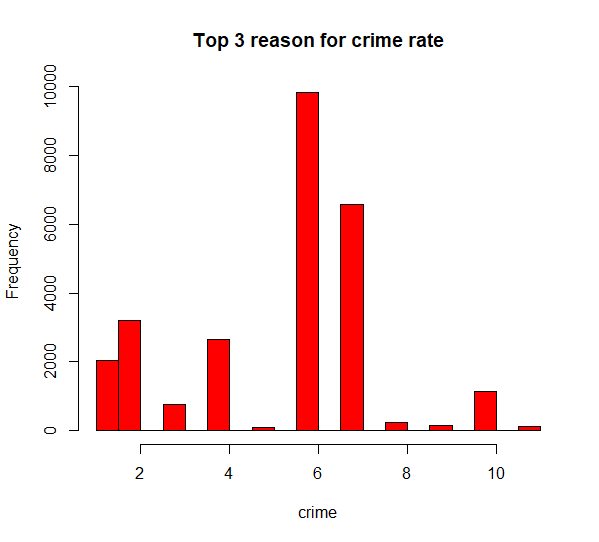
13.1

1. Use the given link Data Set

Answer the below questions:

1. Find out top 5 attributes having highest correlation (select only Numeric features).
2. yeast <- read.csv("C:/Users/satish/Desktop/yeast.txt", sep="")
3. > View(yeast)
4. > View(yeast)
5. > cor1 <- cor.test(yeast$mcg, yeast$gvh, method = "pearson")
6. > cor1
7. Pearson's product-moment correlation
8. data: yeast$mcg and yeast$gvh
9. t = 27.526, df = 1482, p-value < 2.2e-16
10. alternative hypothesis: true correlation is not equal to 0
11. 95 percent confidence interval:
12. 0.5469332 0.6143348
13. sample estimates:
14. cor
15. 0.5816314
16. > cor2<- cor.test(yeast$gvh, yeast$alm, method = "pearson")
17. > cor2
18. Pearson's product-moment correlation
19. data: yeast$gvh and yeast$alm
20. t = -10.873, df = 1482, p-value < 2.2e-16
21. alternative hypothesis: true correlation is not equal to 0
22. 95 percent confidence interval:
23. -0.3182836 -0.2240126
24. sample estimates:
25. cor
26. -0.2718
27. > cor3<- cor.test(yeast$alm, yeast$mit, method = "pearson")
28. > cor3
29. Pearson's product-moment correlation
30. data: yeast$alm and yeast$mit
31. t = 2.3011, df = 1482, p-value = 0.02152
32. alternative hypothesis: true correlation is not equal to 0
33. 95 percent confidence interval:
34. 0.008809399 0.110219289
35. sample estimates:
36. cor
37. 0.0596683
38. > cor4<- cor.test(yeast$mit, yeast$erl, method = "pearson")
39. > cor4
40. Pearson's product-moment correlation
41. data: yeast$mit and yeast$erl
42. t = -0.22832, df = 1482, p-value = 0.8194
43. alternative hypothesis: true correlation is not equal to 0
44. 95 percent confidence interval:
45. -0.05679921 0.04496851
46. sample estimates:
47. cor
48. -0.005930705
49. > cor5<- cor.test(yeast$erl, yeast$mcg, method = "pearson")
50. > cor5
51. Pearson's product-moment correlation
52. data: yeast$erl and yeast$mcg
53. t = 2.5046, df = 1482, p-value = 0.01237
54. alternative hypothesis: true correlation is not equal to 0
55. 95 percent confidence interval:
56. 0.01408241 0.11542587
57. sample estimates:
58. cor
59. 0.06492154

b. Find out top 3 reasons for having more crime in a city.



1. LARCENY-NON VEHICLE 2, LARCENY-VEHICLE 3. AUTO THEFT

c. Which all attributes have correlation with crime rate?

> cor11<- cor.test(COBRA.YTD2017$offense\_id, COBRA.YTD2017$beat, method = "pearson")

> cor11

Pearson's product-moment correlation

data: COBRA.YTD2017$offense\_id and COBRA.YTD2017$beat

t = -3.6751, df = 26757, p-value = 0.0002383

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.03443370 -0.01048248

sample estimates:

cor

-0.02246131

> cor12<- cor.test(COBRA.YTD2017$offense\_id, COBRA.YTD2017$MinOfucr, method = "pearson")

> cor12

Pearson's product-moment correlation

data: COBRA.YTD2017$offense\_id and COBRA.YTD2017$MinOfucr

t = -38.827, df = 26757, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.2422611 -0.2195757

sample estimates:

cor

-0.2309498

> cor13<- cor.test(COBRA.YTD2017$offense\_id, COBRA.YTD2017$MaxOfnum\_victims, method = "pearson")

> cor13

Pearson's product-moment correlation

data: COBRA.YTD2017$offense\_id and COBRA.YTD2017$MaxOfnum\_victims

t = 2.5494, df = 26682, p-value = 0.0108

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.003607442 0.027598557

sample estimates:

cor

0.01560525

> cor14<- cor.test(COBRA.YTD2017$offense\_id, COBRA.YTD2017$loc\_type, method = "pearson")

> cor14

Pearson's product-moment correlation

data: COBRA.YTD2017$offense\_id and COBRA.YTD2017$loc\_type

t = 0.12292, df = 23413, p-value = 0.9022

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.01200549 0.01361189

sample estimates:

cor

0.0008033295