APS Odd Semester 2024 Coding Problem Set-2

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(corrplot)

## corrplot 0.95 loaded

# Load the house price d  
hData = read.csv('houseprices.csv', header = TRUE, stringsAsFactors = FALSE, na.strings = c("", "NA", "Not Available", "not available"))  
str(hData)

## 'data.frame': 225 obs. of 8 variables:  
## $ locality : chr "BTM Layout" "BTM Layout" "BTM Layout" "BTM Layout" ...  
## $ area : int 565 1837 1280 2220 1113 1332 1815 1400 3006 1600 ...  
## $ rent : int 20060 97434 54448 117000 34388 36394 112000 41266 129000 92849 ...  
## $ price\_per\_sqft: int 6195 9254 7422 9234 5391 4767 10744 5143 7485 10125 ...  
## $ facing : chr "North-West" "East" "East" "North" ...  
## $ BHK : int 1 3 2 3 2 2 3 2 4 3 ...  
## $ bathrooms : int 1 3 2 3 2 2 2 2 5 2 ...  
## $ parking : chr "Bike" "Bike and Car" "Car" "Bike and Car" ...

na.omit(hData)

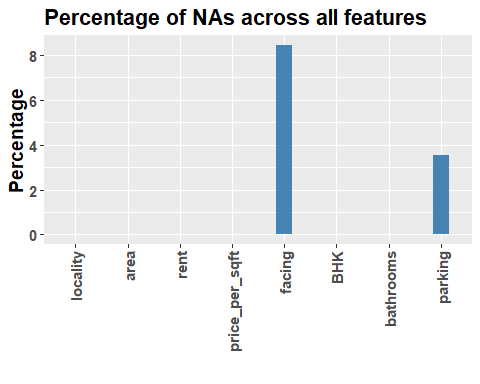
## locality area rent price\_per\_sqft facing BHK bathrooms  
## 1 BTM Layout 565 20060 6195 North-West 1 1  
## 2 BTM Layout 1837 97434 9254 East 3 3  
## 3 BTM Layout 1280 54448 7422 East 2 2  
## 4 BTM Layout 2220 117000 9234 North 3 3  
## 5 BTM Layout 1113 34388 5391 East 2 2  
## 6 BTM Layout 1332 36394 4767 West 2 2  
## 7 BTM Layout 1815 112000 10744 North-East 3 2  
## 8 BTM Layout 1400 41266 5143 South-East 2 2  
## 9 BTM Layout 3006 129000 7485 East 4 5  
## 10 BTM Layout 1600 92849 10125 South 3 2  
## 11 BTM Layout 1200 42985 6250 West 2 2  
## 12 BTM Layout 1000 34388 6000 West 2 2  
## 13 BTM Layout 1200 40120 5833 North 2 2  
## 14 BTM Layout 4650 229000 8602 East 4 13  
## 15 BTM Layout 1060 31522 5189 North 3 2  
## 16 BTM Layout 1325 71643 9434 East 3 2  
## 17 BTM Layout 1025 32096 5463 North 2 2  
## 18 BTM Layout 2200 71643 5682 East 3 3  
## 19 BTM Layout 920 22925 4348 North 2 2  
## 20 BTM Layout 1030 27510 4660 North 2 2  
## 21 BTM Layout 1020 33242 5686 West 2 2  
## 22 BTM Layout 1650 42985 4545 East 3 2  
## 23 BTM Layout 1100 51583 8182 West 2 2  
## 24 BTM Layout 1185 40120 5907 North 3 2  
## 25 BTM Layout 2220 117000 9234 North 3 3  
## 26 Attibele 765 18340 4183 East 1 1  
## 27 Attibele 656 13182 3506 North-East 2 2  
## 28 Attibele 837 18913 3943 East 2 2  
## 29 Attibele 1347 21091 2732 North 3 2  
## 30 Attibele 691 20060 5065 East 2 1  
## 31 Attibele 750 15474 3600 East 2 2  
## 32 Attibele 1069 21206 3461 North 2 2  
## 33 Attibele 1007 25791 4469 East 2 2  
## 34 Attibele 1258 25791 3577 East 2 2  
## 35 Attibele 1169 25791 3849 West 2 2  
## 36 Attibele 1350 26937 3481 North-East 2 2  
## 37 Attibele 1175 25218 3745 East 2 2  
## 38 Attibele 2900 54448 3276 North 4 4  
## 39 Attibele 1060 20633 3396 South 2 2  
## 40 Attibele 1007 31522 5462 East 2 2  
## 41 Attibele 1210 21779 3140 North-East 3 3  
## 42 Attibele 880 16048 3182 North-East 2 2  
## 43 Attibele 810 18340 3951 East 2 2  
## 44 Attibele 800 14328 3125 North 2 2  
## 45 Attibele 750 12609 2933 North 2 1  
## 46 Attibele 2000 13468 1175 East 2 1  
## 47 Attibele 615 17194 4878 East 2 1  
## 48 Attibele 630 18340 5079 East 2 2  
## 49 Attibele 1200 28657 4167 North 3 3  
## 50 Attibele 750 20060 4667 North-East 2 2  
## 51 K R Puram 1164 37254 5584 West 2 2  
## 52 K R Puram 1138 35534 5448 West 2 2  
## 53 K R Puram 1139 36108 5531 West 2 2  
## 54 K R Puram 1139 36108 5531 West 2 2  
## 55 K R Puram 1139 36108 5531 West 2 2  
## 56 K R Puram 1133 36108 5560 West 2 2  
## 57 K R Puram 1228 36108 5130 East 2 2  
## 58 K R Puram 1064 33886 5557 West 2 2  
## 59 K R Puram 1138 35534 5448 West 2 2  
## 60 K R Puram 1295 44203 5956 East 3 2  
## 61 K R Puram 1425 47651 5834 East 3 2  
## 62 K R Puram 1339 41839 5452 East 3 2  
## 63 K R Puram 1075 35735 5800 West 2 2  
## 64 K R Puram 1339 40120 5228 East 3 2  
## 65 K R Puram 1113 37445 5870 West 2 2  
## 66 K R Puram 1139 37254 5707 West 2 2  
## 67 K R Puram 1139 37254 5707 West 2 2  
## 68 K R Puram 1295 43623 5877 East 3 2  
## 69 K R Puram 1151 35534 5387 West 2 2  
## 70 K R Puram 1139 35534 5443 West 2 2  
## 71 K R Puram 1075 34512 5601 West 2 2  
## 72 K R Puram 1113 36229 5679 West 2 2  
## 73 K R Puram 1143 38471 5873 North 2 2  
## 74 K R Puram 1064 35534 5827 West 2 2  
## 75 K R Puram 1143 38523 5881 North 2 2  
## 78 Marathahalli 1550 46997 5290 West 3 3  
## 79 Marathahalli 1120 37254 5804 West 2 2  
## 80 Marathahalli 1560 50150 5609 East 3 3  
## 81 Marathahalli 1560 54448 6090 East 3 3  
## 82 Marathahalli 1110 42412 6667 East 2 2  
## 83 Marathahalli 1050 31522 5238 East 2 2  
## 84 Marathahalli 1500 42985 5000 West 3 3  
## 85 Marathahalli 950 25791 4737 East 2 2  
## 86 Marathahalli 1600 51583 5625 West 3 3  
## 87 Marathahalli 1302 48717 6528 East 2 2  
## 88 Marathahalli 1450 48717 5862 West 3 3  
## 89 Marathahalli 1170 39833 5940 North 2 2  
## 90 Marathahalli 1000 28657 5000 East 2 2  
## 91 Marathahalli 1000 37254 6500 North 2 2  
## 92 Marathahalli 1650 45851 4848 East 3 3  
## 93 Marathahalli 1640 48717 5183 North 3 2  
## 94 Marathahalli 1168 31522 4709 South 2 2  
## 96 Marathahalli 1145 31522 4803 East 2 2  
## 97 Marathahalli 1204 42412 6146 East 2 2  
## 98 Marathahalli 1204 37953 5500 West 2 2  
## 99 Marathahalli 1300 28657 3846 North 2 2  
## 100 Marathahalli 1034 24072 4062 West 2 2  
## 101 Indiranagar 2200 235000 18636 North 3 3  
## 102 Indiranagar 2750 163000 10364 North 4 3  
## 103 Indiranagar 1100 51583 8182 North 2 2  
## 105 Indiranagar 1780 84825 8315 North-West 3 3  
## 106 Indiranagar 1780 91703 8989 South 3 3  
## 107 Indiranagar 1300 65911 8846 East 3 2  
## 108 Indiranagar 1400 40120 5000 North 2 2  
## 109 Indiranagar 925 37254 7027 North 2 2  
## 110 Indiranagar 1475 120000 14237 East 3 2  
## 111 Indiranagar 1200 57314 8333 North-West 2 2  
## 113 Indiranagar 1800 25791 2500 East 4 3  
## 114 Indiranagar 1445 54448 6574 South-East 3 2  
## 115 Indiranagar 1900 103000 9474 East 3 2  
## 118 Indiranagar 1437 81386 9882 North 2 2  
## 119 Indiranagar 1100 48717 7727 North 2 2  
## 120 Indiranagar 2258 115000 8902 East 3 4  
## 121 Indiranagar 1331 48717 6386 East 3 2  
## 122 Indiranagar 2700 126000 8148 South 4 6  
## 123 Indiranagar 1070 71643 11682 South 2 2  
## 124 Indiranagar 1250 65911 9200 South 2 2  
## 125 Indiranagar 1243 56741 7965 South 3 2  
## 126 Electronic City 1200 29803 4333 West 2 2  
## 127 Electronic City 900 22925 4444 West 2 2  
## 129 Electronic City 1395 55595 6953 West 3 2  
## 130 Electronic City 1211 52156 7514 East 2 2  
## 131 Electronic City 1175 38973 5787 East 2 2  
## 132 Electronic City 1170 37254 5556 East 2 2  
## 133 Electronic City 1500 41266 4800 North 3 2  
## 134 Electronic City 1350 37254 4815 West 3 3  
## 135 Electronic City 1470 56741 6735 South 3 3  
## 136 Electronic City 540 10603 3426 North 1 1  
## 137 Electronic City 1270 54448 7480 North 2 2  
## 139 Electronic City 990 40120 7071 East 2 2  
## 140 Electronic City 1558 41266 4621 North 3 3  
## 141 Electronic City 1585 60180 6625 East 3 3  
## 142 Electronic City 1360 29803 3824 West 3 2  
## 143 Electronic City 995 28084 4925 East 2 2  
## 144 Electronic City 965 22925 4145 West 2 2  
## 145 Electronic City 1220 30376 4344 East 3 2  
## 148 Electronic City 1365 60180 7692 North-East 2 2  
## 149 Electronic City 1108 39546 6227 East 2 2  
## 150 Electronic City 1325 37254 4906 North 2 2  
## 152 Yalahanka 1100 28657 4545 East 2 2  
## 153 Yalahanka 1580 57314 6329 West 3 2  
## 154 Yalahanka 550 12609 4000 North 1 1  
## 155 Yalahanka 1275 42985 5882 North 3 2  
## 156 Yalahanka 600 13755 4000 North 1 2  
## 158 Yalahanka 475 11119 4084 West 1 1  
## 159 Yalahanka 485 11749 4227 East 1 1  
## 160 Yalahanka 700 10889 2714 North-East 1 1  
## 162 Yalahanka 1250 39546 5520 North-East 3 2  
## 163 Yalahanka 500 11462 4000 East 1 2  
## 164 Yalahanka 750 25791 6000 North-West 2 2  
## 165 Yalahanka 1500 42985 5000 East 2 2  
## 168 Yalahanka 1350 54448 7037 East 3 2  
## 169 Yalahanka 2661 100000 6576 East 3 3  
## 170 Yalahanka 1096 57314 9124 North 2 2  
## 171 Yalahanka 468 6877 2564 South 1 1  
## 172 Yalahanka 1390 42985 5396 East 2 2  
## 173 Yalahanka 1065 31522 5164 North 2 2  
## 174 Yalahanka 1800 51583 5000 East 4 4  
## 175 Yalahanka 1150 54448 8261 North-East 2 2  
## 177 Malleshwaram 1582 115000 12642 South 3 3  
## 178 Malleshwaram 937 48717 9072 West 2 2  
## 179 Malleshwaram 882 42985 8503 North 2 2  
## 180 Malleshwaram 1850 91703 8649 South 4 4  
## 181 Malleshwaram 1770 129000 12712 East 3 4  
## 183 Malleshwaram 1550 112000 12581 East 3 3  
## 184 Malleshwaram 900 48717 9444 East 2 1  
## 185 Malleshwaram 1150 85971 13043 West 3 2  
## 187 Malleshwaram 965 63045 11399 East 2 2  
## 188 Malleshwaram 1250 68777 9600 North 2 2  
## 189 Malleshwaram 1960 80240 7143 East 3 3  
## 190 Malleshwaram 1960 91703 8163 East 3 3  
## 191 Malleshwaram 1185 54448 8017 East 2 2  
## 192 Malleshwaram 950 54448 10000 North 2 2  
## 193 Malleshwaram 1400 85971 10714 East 3 2  
## 194 Malleshwaram 700 42985 10714 North-East 2 2  
## 195 Malleshwaram 650 25791 6923 South 2 2  
## 196 Malleshwaram 1200 66484 9667 South-East 3 2  
## 197 Malleshwaram 750 48717 11333 South 2 2  
## 198 Malleshwaram 2016 85971 7440 East 4 4  
## 199 Malleshwaram 900 39546 7667 East 2 1  
## 200 Malleshwaram 850 42985 8824 South 2 2  
## 201 Jayanagar 1990 201000 17588 South 3 3  
## 203 Jayanagar 1200 74508 10833 North 2 2  
## 204 Jayanagar 2628 172000 11416 South 3 3  
## 205 Jayanagar 1485 85971 10101 South-East 3 3  
## 209 Jayanagar 2000 109000 9500 East 3 3  
## 210 Jayanagar 1774 132000 12965 West 3 3  
## 212 Jayanagar 1100 53302 8455 North-West 2 2  
## 213 Jayanagar 1100 31522 5000 East 2 3  
## 214 Jayanagar 1410 97434 12057 North 3 3  
## 215 Jayanagar 1940 152000 13660 West 3 3  
## 217 Jayanagar 3223 229000 12411 East 4 4  
## 218 Jayanagar 2495 201000 14028 East 4 4  
## 219 Jayanagar 850 51583 10588 North-East 2 2  
## 220 Jayanagar 1200 42985 6250 East 3 2  
## 222 Jayanagar 2000 85971 7500 East 3 3  
## 223 Jayanagar 1340 77947 10149 East 3 2  
## 224 Jayanagar 5100 430000 14706 East 4 5  
## 225 Jayanagar 1340 74508 9701 East 3 2  
## parking  
## 1 Bike  
## 2 Bike and Car  
## 3 Car  
## 4 Bike and Car  
## 5 Bike and Car  
## 6 Bike and Car  
## 7 Car  
## 8 Bike and Car  
## 9 Bike and Car  
## 10 Bike and Car  
## 11 Bike and Car  
## 12 Bike and Car  
## 13 Bike and Car  
## 14 Bike  
## 15 Bike and Car  
## 16 Bike and Car  
## 17 Bike and Car  
## 18 Bike and Car  
## 19 Bike and Car  
## 20 Bike and Car  
## 21 Car  
## 22 Car  
## 23 Bike and Car  
## 24 Bike and Car  
## 25 Bike and Car  
## 26 Bike and Car  
## 27 Bike and Car  
## 28 Car  
## 29 Bike and Car  
## 30 Bike and Car  
## 31 Bike and Car  
## 32 Bike and Car  
## 33 Bike and Car  
## 34 Bike and Car  
## 35 Bike and Car  
## 36 Car  
## 37 Bike and Car  
## 38 Car  
## 39 Bike and Car  
## 40 Bike and Car  
## 41 Bike and Car  
## 42 Bike and Car  
## 43 Bike and Car  
## 44 Bike and Car  
## 45 Bike and Car  
## 46 Bike and Car  
## 47 Bike and Car  
## 48 Car  
## 49 Bike and Car  
## 50 Bike and Car  
## 51 Bike and Car  
## 52 Bike and Car  
## 53 Bike and Car  
## 54 Bike and Car  
## 55 Bike and Car  
## 56 Bike and Car  
## 57 Bike and Car  
## 58 Bike and Car  
## 59 Bike and Car  
## 60 Bike and Car  
## 61 Bike and Car  
## 62 Bike and Car  
## 63 Bike and Car  
## 64 Bike and Car  
## 65 Bike and Car  
## 66 Bike and Car  
## 67 Bike and Car  
## 68 Bike and Car  
## 69 Bike and Car  
## 70 Bike and Car  
## 71 Bike and Car  
## 72 Bike and Car  
## 73 Bike and Car  
## 74 Bike and Car  
## 75 Bike and Car  
## 78 Bike and Car  
## 79 Bike and Car  
## 80 Bike and Car  
## 81 Bike and Car  
## 82 Bike and Car  
## 83 Car  
## 84 Bike and Car  
## 85 Bike and Car  
## 86 Bike and Car  
## 87 Bike and Car  
## 88 Bike and Car  
## 89 Bike and Car  
## 90 Bike  
## 91 Bike and Car  
## 92 Bike and Car  
## 93 Bike and Car  
## 94 Car  
## 96 Bike and Car  
## 97 Bike and Car  
## 98 Bike and Car  
## 99 Bike  
## 100 Car  
## 101 Bike and Car  
## 102 Bike and Car  
## 103 Bike and Car  
## 105 Bike and Car  
## 106 Bike and Car  
## 107 Bike and Car  
## 108 Bike and Car  
## 109 Car  
## 110 Bike and Car  
## 111 Bike and Car  
## 113 Bike and Car  
## 114 Bike and Car  
## 115 Bike and Car  
## 118 Car  
## 119 Bike and Car  
## 120 Bike and Car  
## 121 Bike and Car  
## 122 Bike and Car  
## 123 Car  
## 124 Bike and Car  
## 125 Bike and Car  
## 126 Car  
## 127 Bike and Car  
## 129 Bike and Car  
## 130 Bike and Car  
## 131 Bike and Car  
## 132 Bike and Car  
## 133 Bike and Car  
## 134 Car  
## 135 Car  
## 136 Bike  
## 137 Bike and Car  
## 139 Car  
## 140 Bike and Car  
## 141 Bike and Car  
## 142 Bike and Car  
## 143 Bike and Car  
## 144 Car  
## 145 Bike and Car  
## 148 Bike and Car  
## 149 Bike and Car  
## 150 Bike and Car  
## 152 Bike and Car  
## 153 Bike and Car  
## 154 Bike and Car  
## 155 Bike and Car  
## 156 Bike and Car  
## 158 Bike  
## 159 Bike and Car  
## 160 Bike and Car  
## 162 Bike and Car  
## 163 Bike and Car  
## 164 Bike  
## 165 Bike and Car  
## 168 Car  
## 169 Car  
## 170 Car  
## 171 Bike  
## 172 Bike and Car  
## 173 Bike and Car  
## 174 Bike  
## 175 Bike and Car  
## 177 Car  
## 178 Bike and Car  
## 179 Bike and Car  
## 180 Bike and Car  
## 181 Bike and Car  
## 183 Bike and Car  
## 184 Bike and Car  
## 185 Bike and Car  
## 187 Car  
## 188 Bike and Car  
## 189 Car  
## 190 Bike and Car  
## 191 Bike and Car  
## 192 Bike and Car  
## 193 Bike and Car  
## 194 Bike and Car  
## 195 Bike and Car  
## 196 Bike and Car  
## 197 Car  
## 198 Car  
## 199 Car  
## 200 Bike and Car  
## 201 Bike and Car  
## 203 Bike and Car  
## 204 Bike and Car  
## 205 Bike and Car  
## 209 Bike and Car  
## 210 Car  
## 212 Bike and Car  
## 213 Bike  
## 214 Bike and Car  
## 215 Bike and Car  
## 217 Car  
## 218 Bike and Car  
## 219 Bike and Car  
## 220 Bike and Car  
## 222 Bike and Car  
## 223 Bike and Car  
## 224 Bike and Car  
## 225 Bike and Car

# Convert 'locality', 'facing' and 'parking' columns to factors  
categorical\_cols = c('locality', 'facing', 'parking')  
hData[categorical\_cols] = lapply(hData[categorical\_cols], as.factor)  
str(hData)

## 'data.frame': 225 obs. of 8 variables:  
## $ locality : Factor w/ 9 levels "Attibele","BTM Layout",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ area : int 565 1837 1280 2220 1113 1332 1815 1400 3006 1600 ...  
## $ rent : int 20060 97434 54448 117000 34388 36394 112000 41266 129000 92849 ...  
## $ price\_per\_sqft: int 6195 9254 7422 9234 5391 4767 10744 5143 7485 10125 ...  
## $ facing : Factor w/ 7 levels "East","North",..: 4 1 1 2 1 7 3 6 1 5 ...  
## $ BHK : int 1 3 2 3 2 2 3 2 4 3 ...  
## $ bathrooms : int 1 3 2 3 2 2 2 2 5 2 ...  
## $ parking : Factor w/ 3 levels "Bike","Bike and Car",..: 1 2 3 2 2 2 3 2 2 2 ...

# Continuous columns  
continuous\_cols = setdiff(colnames(hData), categorical\_cols)

# Plot percentage of NAs in each column of the data frame  
hData\_NA = setNames(stack(sapply(hData, function(x){(sum(is.na(x))/length(x))\*100}))[2:1], c('Feature','Value'))  
p = ggplot(data = hData\_NA, aes(x = Feature, y = Value)) +  
 geom\_bar(stat = 'identity', fill = 'steelblue', width = 0.3) +  
 theme(text = element\_text(size = 14, face = 'bold'),  
 axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5)) +  
 xlab('') + ylab('Percentage') +  
 ggtitle('Percentage of NAs across all features')  
p

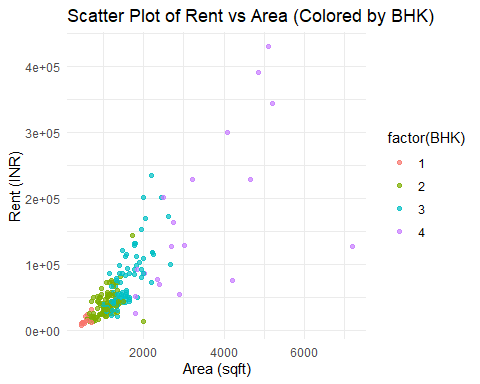


#can see that there are none values in facing and parking

# Add NA as a factor level for categorical columns   
hData[categorical\_cols] = lapply(hData[categorical\_cols], addNA)  
str(hData)

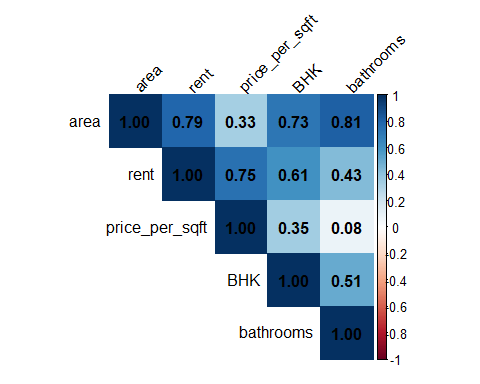
## 'data.frame': 225 obs. of 8 variables:  
## $ locality : Factor w/ 10 levels "Attibele","BTM Layout",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ area : int 565 1837 1280 2220 1113 1332 1815 1400 3006 1600 ...  
## $ rent : int 20060 97434 54448 117000 34388 36394 112000 41266 129000 92849 ...  
## $ price\_per\_sqft: int 6195 9254 7422 9234 5391 4767 10744 5143 7485 10125 ...  
## $ facing : Factor w/ 8 levels "East","North",..: 4 1 1 2 1 7 3 6 1 5 ...  
## $ BHK : int 1 3 2 3 2 2 3 2 4 3 ...  
## $ bathrooms : int 1 3 2 3 2 2 2 2 5 2 ...  
## $ parking : Factor w/ 4 levels "Bike","Bike and Car",..: 1 2 3 2 2 2 3 2 2 2 ...

# Scatter Plot: Rent vs Area, colored by BHK  
ggplot(hData, aes(x = area, y = rent, color = factor(BHK))) +  
 geom\_point(alpha = 0.7) +  
 labs(title = "Scatter Plot of Rent vs Area (Colored by BHK)",  
 x = "Area (sqft)", y = "Rent (INR)") +  
 theme\_minimal()



#1-As the area increases, rent generally tends to rise.  
#2-Properties with higher BHK values are clustered at higher rents, indicating that #larger properties (more rooms) are priced higher.

# correlation matrix  
numeric\_cols <- hData[, c("area", "rent", "price\_per\_sqft", "BHK", "bathrooms")]  
  
# Compute the correlation matrix  
cor\_matrix <- cor(numeric\_cols, use = "complete.obs")  
corrplot(cor\_matrix, method = "color", type = "upper",   
 tl.col = "black", tl.srt = 45, addCoef.col = "black")

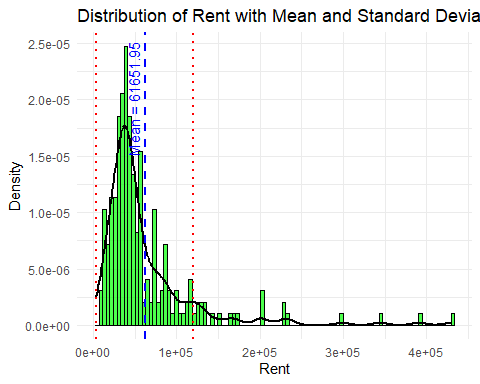


#Correlation Analysis -   
 #1- (r=0.79): Rent tends to increase as area increases.  
   
#2- Price per Sqft and Bathrooms (0.080264)-  
#very weak positive correlation btw price\_per\_sqft and bathrooms. The number of bathrooms has little to no effect on the price per square foot.  
  
#3- Area and Bathrooms (0.812822):  
#There is a strong positive correlation between area and bathrooms. Larger properties tend to have more bathrooms.  
  
#4-Price per Sqft and Bathrooms (0.080264):  
  
#There is a very weak positive correlation between price\_per\_sqft and bathrooms. The number of bathrooms has little to no effect on the price per square foot.

mean\_rent <- mean(hData$rent)  
sd\_rent <- sd(hData$rent)  
  
# Plot the histogram with density and lines for mean and standard deviation  
ggplot(hData, aes(x = rent)) +  
 geom\_histogram(aes(y = ..density..), bins = 100, fill = "green", color = "black", alpha = 0.7) + # Histogram with outlines  
 geom\_density(color = "black", size = 1) + # Add density curve  
 geom\_vline(aes(xintercept = mean\_rent), color = "blue", linetype = "dashed", size = 1) + # Mean line  
 geom\_vline(aes(xintercept = mean\_rent + sd\_rent), color = "red", linetype = "dotted", size = 1) + # +1 Standard Deviation  
 geom\_vline(aes(xintercept = mean\_rent - sd\_rent), color = "red", linetype = "dotted", size = 1) + # -1 Standard Deviation  
 labs(title = "Distribution of Rent with Mean and Standard Deviation", x = "Rent", y = "Density") +  
   
 # Add text label for mean rent on the graph  
 annotate("text", x = mean\_rent, y = 0.00002, label = paste("Mean =", round(mean\_rent, 2)), color = "blue", angle = 90, vjust = -0.5, size = 4) +  
   
 theme\_minimal()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.  
## ℹ Please use `after\_stat(density)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

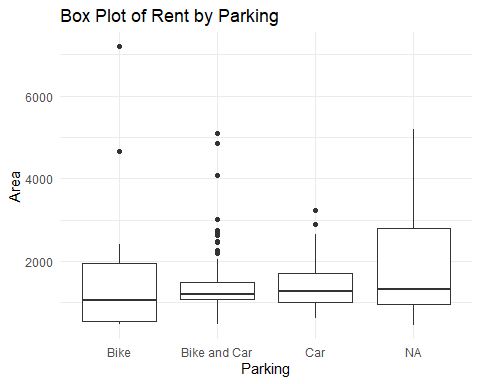


|  |
| --- |
| histogram of rent prices shows a right-skewed distribution, indicating that most properties have relatively lower rent prices, with a few high-rent outliers pulling the average up. The mean rent is approx 61,652, as represented by a dashed blue line on the graph. This wide standard deviation reflects a large variation in rent values, with a spread from affordable to luxury properties |

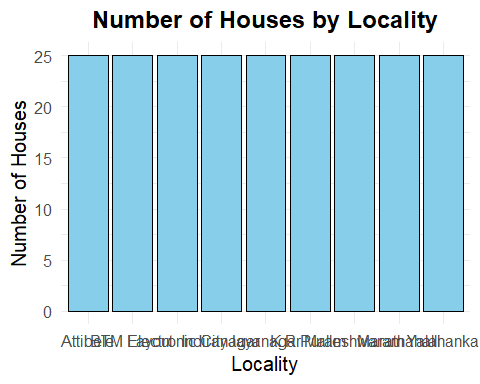
# Calculate and print summary statistics for Rent  
rent\_mean <- mean(hData$rent, na.rm = TRUE)  
rent\_sd <- sd(hData$rent, na.rm = TRUE)  
rent\_var <- var(hData$rent, na.rm = TRUE)  
cat("Rent - Mean:", rent\_mean, " | SD:", rent\_sd, " | Variance:", rent\_var, "\n")

## Rent - Mean: 61651.95 | SD: 58728.55 | Variance: 3449042972

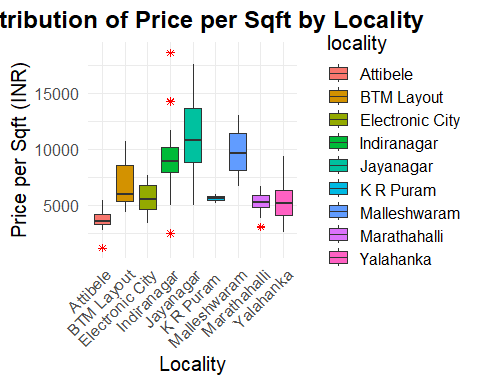
ggplot(hData, aes(x = parking, y = area)) +  
 geom\_boxplot() +  
 labs(title = "Box Plot of Rent by Parking", x = "Parking", y = "Area") +  
 theme\_minimal()



ggplot(hData, aes(x = locality)) +  
 geom\_bar(fill = "skyblue", color = "black") +  
 labs(  
 title = "Number of Houses by Locality",   
 x = "Locality",   
 y = "Number of Houses"  
 ) +  
 theme\_minimal(base\_size = 15) +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold")  
 )



ggplot(hData, aes(x = locality, y = price\_per\_sqft, fill = locality)) +  
 geom\_boxplot(outlier.color = "red", outlier.shape = 8) +  
 labs(  
 title = "Distribution of Price per Sqft by Locality",   
 x = "Locality",   
 y = "Price per Sqft (INR)"  
 ) +  
 theme\_minimal(base\_size = 15) +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 axis.text.x = element\_text(angle = 45, hjust = 1)  
 )



# Box plot of Rent by Parking with colors  
ggplot(hData, aes(x = as.factor(parking), y = rent, fill = as.factor(parking))) +  
 geom\_boxplot() +  
 labs(title = "Rent Distribution by Parking Spaces",  
 x = "Parking Spaces",  
 y = "Rent (INR)") +  
 scale\_fill\_brewer(palette = "Set3") + # Choose a color palette  
 theme\_minimal() +  
 theme(legend.title = element\_blank()) # Remove legend title

