#Problem 2

#Calculate the Manhattan distance between any two vectors, vec1 and vec2

#Assign vectors, fill with any two vectors #vec1 <- c(1,2,3,4) #vec2 <- c(2,3,4,5)

Manhattan_dist <- function(vec1, vec2){ dist <- sum(abs(vec1 - vec2)) return (dist) }

Manhattan_dist(vec1,vec2)

#Calculate the Euclidian distance between any two vectors, vec1 and vec2

#Assign vectors, fill with any two vectors #vec1 <- c(1,2,3,4) #vec2 <- c(2,3,4,5)

Euclidian_dist <- function(vec1, vec2){ dist <- sqrt(sum((vec1 - vec2)^2)) return (dist) }

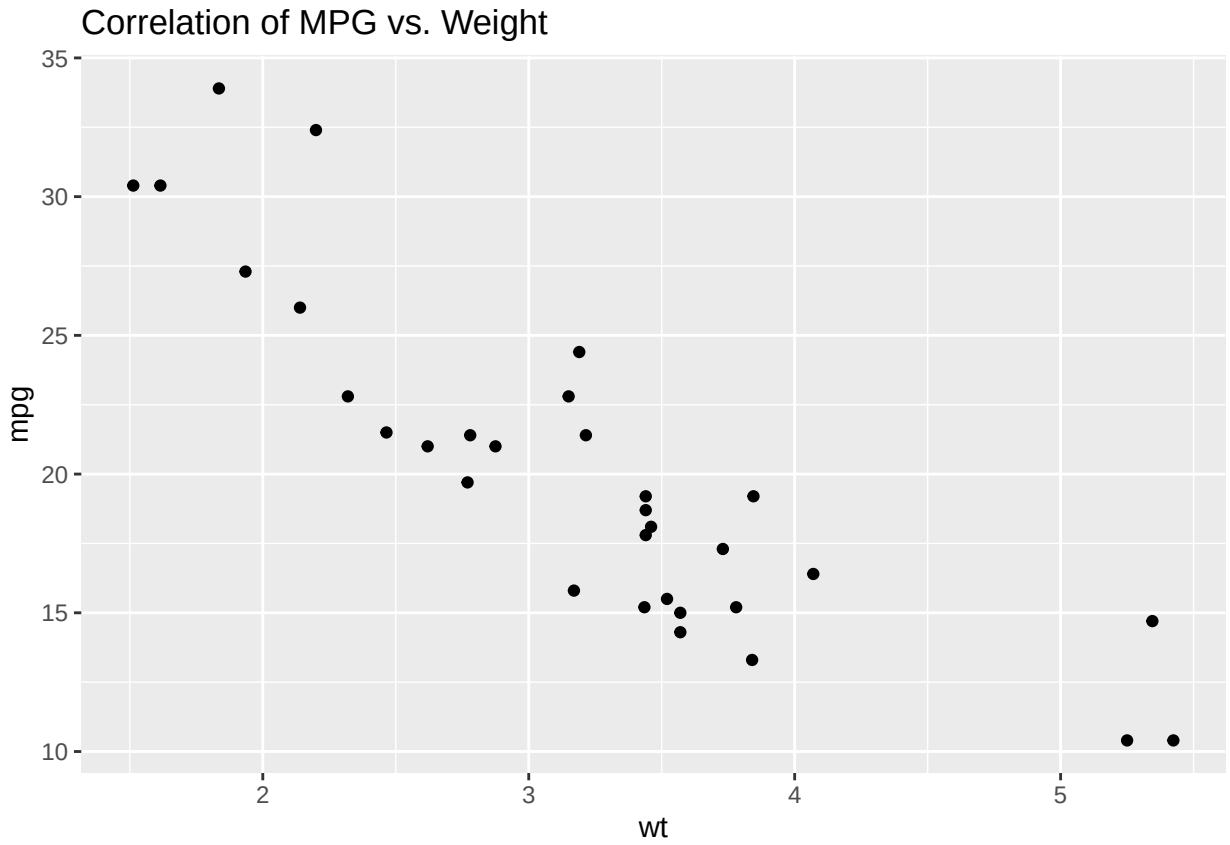
Euclidian_dist(vec1,vec2)

```r
#Problem #4

library(ggplot2)
data("mtcars")

#Calculate correlation between mpg and wt
cor_mpg_wt <- cor(mtcars$mpg, mtcars$wt)

#Plot correlation onto scatter plot
ggplot(mtcars, aes(x = wt, y = mpg)) +
  geom_point() +
  labs(title = "Correlation of MPG vs. Weight")
```



```r
#Problem #5

setwd("C:/Users/kathryn.poniatowski/OneDrive - Mass Transport Authority, Flint/UofM/CSC 587/Poniatowski

#Import test data into dataframe
metabolite <- read.csv("metabolite.csv")

#Calculate missing precentage in columns
missing_percentage <- colMeans(is.na(metabolite))

#Drop columns with over 75% missing
metabolite_columns <- missing_percentage <= 0.75

#Assign columns to amended dataframe
metabolite_amended <- metabolite[, metabolite_columns]

# Replace missing values in remaining columns with the median of the column
metabolite_amended <- metabolite_amended
for (col in colnames(metabolite_amended)) {
  # Calculate the median excluding NA values
  median_value <- median(metabolite_amended[[col]], na.rm = TRUE)

  # Replace NA values with the median
  metabolite_amended[[col]][is.na(metabolite_amended[[col]])] <- median_value
}

head(metabolite_amended)
```

```
##       Label  Phe Pro Ser Thr ADMA alpha.AAA c4.OH.Pro Carnosine Creatinine
## 1 Alzheimer 72.8 166 170 282 1.15     0.760     0.236     1.270       49.9
## 2 Alzheimer 93.4 138 142 217 1.05     0.929     0.189     1.350       48.8
## 3 Alzheimer 68.6 161 158 208 1.00     0.620     0.198     0.998       30.4
## 4 Alzheimer 94.1 129 162 201 1.10     0.795     0.198     0.675       80.1
## 5 Alzheimer 79.8 126 115 199 1.24     1.360     0.198     1.280       60.5
## 6 Alzheimer 82.5 167 173 333 1.35     1.150     0.198     1.010       24.0
##    DOPA Dopamine Histamine Kynurenine Met.SO Putrescine Sarcosine Serotonin
## 1 0.265    0.233     0.225       5.21  0.526      0.068      17.8     0.147
## 2 0.252    0.231     0.211       5.44  0.387      0.087      20.2     0.231
## 3 0.268    0.231     0.217       5.20  0.651      0.260      14.4     0.196
## 4 0.264    0.234     0.209       5.80  0.389      0.110      18.7     0.255
## 5 0.271    0.231     0.210       4.46  0.466      0.118      22.5     0.390
## 6 0.275    0.231     0.212       7.01  0.417      0.262      30.8     0.140
##   Spermidine t4.OH.Pro Taurine SDMA   C0   C10 C10.1 C10.2   C12 C12.DC C12.1
## 1      0.188      24.0     125 1.13 18.2 0.059 0.312 0.038 0.030  0.042 0.290
## 2      0.233      29.3     120 1.65 17.0 0.051 0.288 0.039 0.038  0.038 0.265
## 3      0.384      20.9     139 1.57 12.6 0.083 0.357 0.054 0.032  0.048 0.302
## 4      0.353      23.1     159 1.34 23.5 0.071 0.317 0.040 0.045  0.048 0.275
## 5      0.473      26.9     149 1.24 13.6 0.139 0.472 0.074 0.056  0.079 0.394
## 6      0.856      26.0     379 1.44 26.7 0.058 0.238 0.042 0.039  0.035 0.196
##     C14 C14.1 C14.1.OH C14.2 C14.2.OH   C16 C16.OH C16.1 C16.1.OH C16.2
## 1 0.023 0.019    0.008 0.008    0.006 0.046  0.008 0.009    0.007 0.005
## 2 0.026 0.017    0.008 0.009    0.009 0.070  0.009 0.013    0.006 0.006
## 3 0.021 0.031    0.010 0.010    0.009 0.076  0.011 0.019    0.010 0.005
## 4 0.026 0.028    0.010 0.013    0.011 0.074  0.011 0.015    0.008 0.006
```

```
## 5 0.034 0.043    0.016 0.025    0.017 0.062  0.007 0.024    0.014 0.012
## 6 0.029 0.023    0.009 0.010    0.007 0.081  0.006 0.012    0.005 0.007
##   C16.2.OH   C18 C18.1 C18.1.OH C18.2   C2    C3 C3.OH  C3.1    C4
## 1    0.013 0.013 0.024    0.003 0.016 1.97 0.354 0.008 0.015 0.082
## 2    0.012 0.014 0.025    0.003 0.028 1.95 0.184 0.009 0.013 0.108
## 3    0.013 0.016 0.025    0.004 0.018 1.70 0.371 0.011 0.012 0.057
## 4    0.009 0.020 0.035    0.004 0.033 2.10 0.278 0.010 0.017 0.110
## 5    0.025 0.031 0.034    0.012 0.017 5.62 0.436 0.029 0.035 0.106
## 6    0.015 0.017 0.035    0.004 0.029 3.49 0.461 0.008 0.014 0.123
##   C3.DC..C4.OH.  C4.1    C5 C5.M.DC C5.OH..C3.DC.M.  C5.1 C5.1.DC C6..C4.1.DC.
## 1         0.045 0.025 0.094   0.023           0.026 0.030   0.020        0.022
## 2         0.080 0.025 0.077   0.032           0.026 0.024   0.021        0.030
## 3         0.035 0.039 0.096   0.045           0.024 0.037   0.018        0.022
## 4         0.077 0.031 0.145   0.034           0.041 0.035   0.016        0.029
## 5         0.099 0.069 0.141   0.094           0.058 0.073   0.049        0.052
## 6         0.068 0.026 0.090   0.019           0.037 0.022   0.016        0.063
##   C5.DC..C6.OH.  C6.1 C7.DC    C8    C9 lysoPC.a.C14.0 lysoPC.a.C16.0
## 1         0.014 0.018 0.011 0.062 0.016           2.23           37.9
## 2         0.018 0.015 0.010 0.058 0.014           1.97           22.1
## 3         0.029 0.031 0.021 0.090 0.017           2.12           33.7
## 4         0.016 0.027 0.017 0.091 0.018           2.19           32.8
## 5         0.040 0.040 0.036 0.192 0.041           1.88           24.5
## 6         0.016 0.019 0.014 0.073 0.014           2.11           29.1
##   lysoPC.a.C16.1 lysoPC.a.C17.0 lysoPC.a.C18.0 lysoPC.a.C18.1 lysoPC.a.C18.2
## 1           2.66          0.446           9.00           8.58           7.27
## 2           1.31          0.270           5.35           3.94           4.42
## 3           2.53          0.399           7.51           7.73           8.02
## 4           2.39          0.323           7.21           7.22           7.62
## 5           1.27          0.382           6.66           5.39           3.60
## 6           2.09          0.348           5.84           6.30           8.10
##   lysoPC.a.C20.3 lysoPC.a.C20.4 lysoPC.a.C24.0 lysoPC.a.C26.0 lysoPC.a.C26.1
## 1          1.830           8.25          0.079          0.113          0.053
## 2          0.958           4.60          0.059          0.066          0.042
## 3          2.050           9.84          0.075          0.126          0.049
## 4          1.640           6.75          0.066          0.086          0.045
## 5          0.970           6.26          0.084          0.118          0.053
## 6          1.970           7.04          0.083          0.112          0.050
##   lysoPC.a.C28.0 lysoPC.a.C28.1 PC.aa.C24.0 PC.aa.C26.0 PC.aa.C28.1 PC.aa.C30.0
## 1          0.108          0.072       0.082       0.438       0.571        2.35
## 2          0.076          0.058       0.065       0.409       0.521        1.99
## 3          0.078          0.092       0.099       0.458       0.605        2.69
## 4          0.076          0.076       0.076       0.486       0.685        3.33
## 5          0.092          0.072       0.069       0.401       0.513        1.78
## 6          0.099          0.083       0.073       0.450       0.620        2.61
##   PC.aa.C32.0 PC.aa.C32.1 PC.aa.C32.2 PC.aa.C32.3 PC.aa.C34.1 PC.aa.C34.2
## 1        11.4        9.22       0.117       0.092       109.0        71.0
## 2        12.7        5.40       0.117       0.067        64.2        60.5
## 3        16.6       11.60       0.117       0.105       108.0        83.1
## 4        18.6       13.30       0.053       0.079       106.0        93.6
## 5        13.8        5.03       0.117       0.102        83.4        35.9
## 6        14.7        8.98       0.117       0.107        90.2        85.6
##   PC.aa.C34.3 PC.aa.C34.4 PC.aa.C36.0 PC.aa.C36.1 PC.aa.C36.2 PC.aa.C36.3
## 1       1.430       0.200        2.38        21.7        42.4        42.7
## 2       0.879       0.127        2.05        14.3        35.6        24.3
```

```
## 3       1.930       0.210        2.30        19.9        44.9        43.9
## 4       1.590       0.190        2.57        20.9        48.8        41.2
## 5       0.709       0.135        1.83        20.5        28.5        21.9
## 6       1.790       0.213        2.48        15.5        43.2        46.0
##   PC.aa.C36.4 PC.aa.C36.5 PC.aa.C36.6 PC.aa.C38.0 PC.aa.C38.3 PC.aa.C38.4
## 1       120.0        1.86       0.084       1.230        32.1        95.1
## 2        83.7        1.05       0.046       0.946        21.9        78.9
## 3       146.0        2.09       0.057       1.210        34.5       107.0
## 4       122.0        1.76       0.070       1.160        28.7        92.7
## 5        98.1        1.70       0.048       1.100        23.3       101.0
## 6       114.0        3.47       0.103       1.390        28.9        78.0
##   PC.aa.C38.5 PC.aa.C38.6 PC.aa.C40.1 PC.aa.C40.2 PC.aa.C40.3 PC.aa.C40.4
## 1       16.80        41.6       0.195       0.074       0.491        3.48
## 2        9.91        25.1       0.211       0.057       0.358        3.39
## 3       17.50        36.6       0.212       0.118       0.395        3.56
## 4       14.30        29.9       0.220       0.097       0.433        3.59
## 5       13.80        36.2       0.165       0.044       0.525        3.37
## 6       13.10        48.4       0.205       0.120       0.346        2.63
##   PC.aa.C40.5 PC.aa.C40.6 PC.aa.C42.0 PC.aa.C42.1 PC.aa.C42.2 PC.aa.C42.4
## 1        5.66        21.8       0.364       0.226       0.108       0.272
## 2        4.08        14.2       0.419       0.216       0.109       0.336
## 3        5.34        16.7       0.476       0.281       0.118       0.300
## 4        5.06        14.0       0.427       0.223       0.119       0.268
## 5        5.29        22.5       0.125       0.095       0.083       0.206
## 6        3.25        18.9       0.451       0.233       0.135       0.228
##   PC.aa.C42.5 PC.aa.C42.6 PC.ae.C30.0 PC.ae.C30.1 PC.ae.C30.2 PC.ae.C32.1
## 1       0.272       0.291       0.173       0.027       0.022        1.65
## 2       0.317       0.248       0.147       0.024       0.020        2.01
## 3       0.206       0.267       0.209       0.046       0.030        2.40
## 4       0.267       0.254       0.223       0.049       0.023        2.47
## 5       0.205       0.280       0.095       0.082       0.023        1.72
## 6       0.254       0.271       0.221       0.039       0.029        2.01
##   PC.ae.C32.2 PC.ae.C34.0 PC.ae.C34.1 PC.ae.C34.2 PC.ae.C34.3 PC.ae.C36.0
## 1       0.371       0.880        3.66        2.48       0.813       0.498
## 2       0.360       0.763        2.68        2.32       0.905       0.398
## 3       0.477       0.938        4.04        2.95       1.030       0.554
## 4       0.459       0.964        4.06        3.09       1.020       0.552
## 5       0.316       1.060        3.28        1.70       0.722       0.553
## 6       0.397       0.920        3.26        2.58       1.000       0.443
##   PC.ae.C36.1 PC.ae.C36.2 PC.ae.C36.3 PC.ae.C36.4 PC.ae.C36.5 PC.ae.C38.0
## 1        5.64        1.90       1.170        6.96        4.79       0.474
## 2        3.89        1.54       0.873        6.40        5.36       0.325
## 3        5.95        2.29       1.240        9.05        6.63       0.478
## 4        4.75        2.01       1.350        8.36        5.97       0.397
## 5        5.95        1.47       0.760        4.78        4.00       0.430
## 6        4.95        2.05       1.170        7.04        4.47       0.590
##   PC.ae.C38.2 PC.ae.C38.3 PC.ae.C38.4 PC.ae.C38.5 PC.ae.C38.6 PC.ae.C40.1
## 1       0.538        2.66        6.33        5.51        1.95       0.574
## 2       0.127        1.80        5.37        4.49        1.63       0.281
## 3       0.154        2.87        7.06        5.64        1.98       0.759
## 4       0.144        1.97        5.99        5.63        1.97       0.425
## 5       0.246        1.80        5.45        4.34        1.51       0.430
## 6       0.312        2.46        5.55        4.60        1.80       0.481
##   PC.ae.C40.2 PC.ae.C40.3 PC.ae.C40.4 PC.ae.C40.5 PC.ae.C40.6 PC.ae.C42.0
```

```
## 1       0.575       0.940        1.76        1.77        1.59       0.629
## 2       0.491       0.702        1.43        1.55        1.20       0.616
## 3       0.654       0.817        1.51        1.64        1.49       0.686
## 4       0.540       0.742        1.45        1.62        1.25       0.637
## 5       0.432       0.632        1.10        1.25        1.47       0.660
## 6       0.598       0.826        1.25        1.38        1.61       0.669
##   PC.ae.C42.1 PC.ae.C42.2 PC.ae.C42.3 PC.ae.C42.4 PC.ae.C42.5 PC.ae.C44.3
## 1       0.316       0.192       0.277       0.264       0.888       0.065
## 2       0.260       0.157       0.200       0.311       0.840       0.071
## 3       0.356       0.241       0.288       0.319       0.957       0.065
## 4       0.299       0.159       0.208       0.392       0.863       0.069
## 5       0.355       0.138       0.174       0.162       0.513       0.081
## 6       0.265       0.195       0.253       0.316       0.814       0.085
##   PC.ae.C44.4 PC.ae.C44.5 PC.ae.C44.6 SM..OH..C14.1 SM..OH..C16.1 SM..OH..C22.1
## 1       0.168       0.536       0.494         1.420          1.33          2.07
## 2       0.220       0.470       0.515         1.390          1.25          2.47
## 3       0.228       0.565       0.603         1.840          1.58          2.69
## 4       0.237       0.517       0.611         1.720          1.48          2.97
## 5       0.154       0.178       0.134         0.987          1.48          1.96
## 6       0.232       0.554       0.539         1.320          1.12          2.51
##   SM..OH..C22.2 SM..OH..C24.1 SM.C16.0 SM.C16.1 SM.C18.0 SM.C18.1 SM.C20.2
## 1          1.86         0.597     44.9     7.99     14.5    10.40    0.290
## 2          2.20         0.640     42.1     6.88     12.7     8.52    0.211
## 3          2.63         0.665     44.8     8.91     14.6    11.60    0.304
## 4          2.84         0.682     52.4     8.61     17.2    11.50    0.261
## 5          1.74         0.478     40.6     5.86     13.0     8.34    0.196
## 6          2.16         0.640     42.6     8.49     13.0    10.60    0.270
##   SM.C24.0 SM.C24.1 SM.C26.0 SM.C26.1 H1_1   H1 Urea_N L.Arginine_N L.Leucine_N
## 1    12.20     27.3    0.147    0.337 3356 3356 185.05         45.1       55.75
## 2    10.40     25.6    0.130    0.317 2509 2509 201.90         22.5       35.30
## 3    11.50     28.8    0.163    0.364 2661 2661 193.30         21.0       25.40
## 4    11.80     27.9    0.138    0.353 2652 2652 500.80         16.0       27.10
## 5     9.29     20.5    0.111    0.283 2258 2258 132.50         13.2       57.90
## 6     9.58     23.7    0.135    0.316 3031 3031 193.30         32.2       26.50
##   EDTAca_N X2.Hydroxybutyrate X3.Hydroxybutyrate Acetate Acetoacetate Acetone
## 1      2.9              19.80               44.1    20.2         21.4   10.15
## 2      2.0              12.40                8.5    13.2          5.7    5.10
## 3      1.8              11.33               11.7     5.8          9.3    5.60
## 4      2.5              12.70                7.2     9.8          4.8    4.00
## 5      2.5              35.20               44.7    20.2         18.9   18.90
## 6      0.0              17.20               16.0    23.6          7.8    5.50
##   Betaine Carnitine Choline Creatine Dimethyl.sulfone Ethanol Formate Glucose
## 1   32.25      13.1   22.15     26.7             3.55     7.2    28.9 2239.35
## 2   22.00       8.7   14.20     14.5             4.70    16.6    24.6 1489.70
## 3   19.10      15.3   14.50     17.8             2.10     8.1    27.4 1343.90
## 4   13.90       7.7   11.80     14.7             1.30     6.4    14.4  629.50
## 5   33.90      18.5   27.70     35.4             5.50    13.0    40.0 1618.00
## 6   16.90      16.7   25.90     18.6             3.40     5.0    35.5 1791.80
##   Glycerol Hypoxanthine Isobutyrate Isopropanol Lactate Malonate
## 1    449.1         7.35         4.6         3.3  1768.7    11.35
## 2    324.6         6.30         3.6         1.9  1171.6    10.40
## 3    201.3         6.00         2.5         2.5  1938.1    13.10
## 4    322.0         8.60         2.5         4.4  1037.7     7.60
## 5    271.6         0.00         6.1        11.2  2199.9    11.70
```

```
## 6    274.2         8.80         2.3         2.4  1486.7    11.80
```

```r
library(ggplot2)
library(stats)

# Exclude non-numeric columns from the dataframe
numeric_data <- metabolite_amended[sapply(metabolite_amended, is.numeric)]

# Perform PCA on the numeric data (excluding the Label column if it exists)
pca <- prcomp(numeric_data, scale. = TRUE)

# Get the first two principal components (PC1, PC2)
pca_data <- as.data.frame(pca$x[, 1:2])

# Add the 'Label' column from the original dataframe to the PCA data
pca_data$Label <- metabolite_amended$Label

# Plotting the scatter plot of PC1 vs PC2, colored by 'Label'
ggplot(pca_data, aes(x = PC1, y = PC2, color = Label)) +
  geom_point() +
  labs(title = "Metabolite PCA Scatter Plot") +
  scale_color_manual(values = c("Group1" = "blue", "Group2" = "red"))
```

```
## Warning: No shared levels found between `names(values)` of the manual scale and the
## data's colour values.
## No shared levels found between `names(values)` of the manual scale and the
## data's colour values.
```