Model evaluation project

Katarzyna Tokarczuk

2024-05-22

Part 1

Task 1

```
dane <- read.csv2("C:/Users/kasia/Desktop/project-git/weather.csv")
```

Task 2

```
str(dane)
```

```
  'data.frame': 599 obs. of  22 variables:
   $ id                    : int  0 1 2 6 8 9 10 12 14 18 ...
   $ DWD_ID                : int  1 3 44 71 73 78 91 98 116 132 ...
   $ STATION.NAME          : chr  "Aach" "Aachen" "Gro\xdfenkneten" "Albstadt-Badkap" ...
   $ FEDERAL.STATE         : chr  "Baden-W\xfcrttemberg" "Nordrhein-Westfalen" "Niedersachsen" "Baden-W
   $ LAT                   : num  47.8 50.8 52.9 48.2 48.6 ...
   $ LON                   : num  8.85 6.09 8.24 8.98 13.05 ...
   $ ALTITUDE              : num  478 202 44 759 340 65 300 780 213 750 ...
   $ PERIOD                : chr  "1931-1986" "1851-2011" "1971-2016" "1986-2016" ...
   $ RECORD.LENGTH         : int  55 160 45 30 64 55 38 67 67 33 ...
   $ MEAN.ANNUAL.AIR.TEMP  : num  8.2 9.8 9.2 7.4 8.4 9.3 8.2 5.1 8.4 5.7 ...
   $ MEAN.MONTHLY.MAX.TEMP : num  13.1 13.6 13.2 12.2 13.4 13.4 12.7 8.9 12.9 9.2 ...
   $ MEAN.MONTHLY.MIN.TEMP : num  3.5 6.3 5.4 3.3 3.9 5.2 4.1 2.2 4.2 2.7 ...
   $ MEAN.ANNUAL.WIND.SPEED: num  2 3 2 2 1 2 3 3 2 3 ...
   $ MEAN.CLOUD.COVER      : num  67 67 67 66 65 67 72 72 66 64 ...
   $ MEAN.ANNUAL.SUNSHINE  : num  NA 1531 1459 1725 1595 ...
   $ MEAN.ANNUAL.RAINFALL  : num  755 820 759 919 790 794 657 NA NA 915 ...
   $ MAX.MONTHLY.WIND.SPEED: num  2 3 3 2 2 2 3 4 3 3 ...
   $ MAX.AIR.TEMP          : num  32.5 32.3 32.4 30.2 33 32.2 31.6 27.6 33.2 29 ...
   $ MAX.WIND.SPEED        : num  NA 30.2 29.9 NA NA NA NA NA NA NA ...
   $ MAX.RAINFALL          : num  39 36 32 43 43 33 37 NA NA 40 ...
   $ MIN.AIR.TEMP          : num  -16.3 -10.9 -12.6 -15.5 -19.2 -13.3 -15.2 -15.7 -17.5 -17.2 ...
   $ MEAN.RANGE.AIR.TEMP   : num  9.6 7.3 7.8 8.9 9.5 8.2 8.6 6.7 8.6 6.5 ...
```

```
summary(dane)
```

```
      id             DWD_ID      STATION.NAME       FEDERAL.STATE
Min.   :   0.0   Min.   :    1   Length:599         Length:599
1st Qu.: 259.5   1st Qu.: 1368   Class :character   Class :character
Median : 479.0   Median : 2812   Mode  :character   Mode  :character
Mean   : 489.2   Mean   : 2902
3rd Qu.: 731.5   3rd Qu.: 4338
Max.   :1058.0   Max.   :15526

     LAT             LON            ALTITUDE         PERIOD
Min.   :47.40   Min.   : 6.094   Min.   :   1.0   Length:599
1st Qu.:49.27   1st Qu.: 8.477   1st Qu.:  75.0   Class :character
Median :50.64   Median : 9.966   Median : 224.0   Mode  :character
Mean   :50.75   Mean   :10.120   Mean   : 285.3
3rd Qu.:51.96   3rd Qu.:11.703   3rd Qu.: 418.0
Max.   :55.01   Max.   :14.951   Max.   :2964.0

RECORD.LENGTH    MEAN.ANNUAL.AIR.TEMP MEAN.MONTHLY.MAX.TEMP
Min.   : 30.00   Min.   : 2.500       Min.   : 3.30
1st Qu.: 54.00   1st Qu.: 8.000       1st Qu.:12.10
Median : 70.00   Median : 8.500       Median :12.90
Mean   : 80.07   Mean   : 8.401       Mean   :12.66
3rd Qu.:103.00   3rd Qu.: 9.100       3rd Qu.:13.50
Max.   :297.00   Max.   :11.000       Max.   :15.60
                 NA's   :1            NA's   :2
MEAN.MONTHLY.MIN.TEMP MEAN.ANNUAL.WIND.SPEED MEAN.CLOUD.COVER
Min.   :0.300         Min.   :1.000          Min.   :56.0
1st Qu.:3.800         1st Qu.:2.000          1st Qu.:65.0
Median :4.600         Median :2.000          Median :67.0
Mean   :4.488         Mean   :2.124          Mean   :66.8
3rd Qu.:5.300         3rd Qu.:2.000          3rd Qu.:69.0
Max.   :7.300         Max.   :6.000          Max.   :79.0
NA's   :4             NA's   :11             NA's   :11
MEAN.ANNUAL.SUNSHINE MEAN.ANNUAL.RAINFALL MAX.MONTHLY.WIND.SPEED
Min.   :   0         Min.   : 446.0       Min.   :1.000
1st Qu.:1441         1st Qu.: 640.2       1st Qu.:2.000
Median :1543         Median : 737.5       Median :3.000
Mean   :1517         Mean   : 787.2       Mean   :2.721
3rd Qu.:1635         3rd Qu.: 857.0       3rd Qu.:3.000
Max.   :1846         Max.   :1995.0       Max.   :7.000
NA's   :193          NA's   :13           NA's   :11
 MAX.AIR.TEMP   MAX.WIND.SPEED   MAX.RAINFALL    MIN.AIR.TEMP
Min.   :13.90   Min.   : 3.80   Min.   :25.00   Min.   :-25.40
1st Qu.:31.10   1st Qu.:25.45   1st Qu.:34.00   1st Qu.:-16.70
Median :32.20   Median :27.50   Median :36.00   Median :-14.90
Mean   :31.84   Mean   :27.56   Mean   :38.55   Mean   :-14.93
3rd Qu.:33.10   3rd Qu.:29.50   3rd Qu.:41.00   3rd Qu.:-13.30
Max.   :35.40   Max.   :54.30   Max.   :76.00   Max.   : -5.30
NA's   :2       NA's   :380     NA's   :14      NA's   :2
MEAN.RANGE.AIR.TEMP
Min.   : 0.000
1st Qu.: 7.600
Median : 8.400
Mean   : 8.168
3rd Qu.: 8.900
```

```
   Max.   :11.100
```

Task 3

```
#Variables that may be irrelevant for the model predicting annual average precipitation: id, DWD_ID, ST
dane <- subset(dane, select = -c(id, DWD_ID, STATION.NAME, FEDERAL.STATE,PERIOD))
```

Task 4

```
rows <- nrow(dane)
dane <- na.omit(dane)
rows_removed <-rows - nrow(dane)
rows_removed
```

```
  [1] 395
```

Task 5

```
dim(dane)
```

```
  [1] 204  17
```

Task 6

```
library(caTools)
```

```
  Warning: pakiet 'caTools' został zbudowany w wersji R 4.3.3
```

```
split <- sample.split(dane, SplitRatio = 0.7,group=NULL)
train_data <- subset(dane, split == TRUE)
test_data <- subset(dane, split == FALSE)
```

Task 7

```
zm_ob_train = as.vector(train_data['MEAN.ANNUAL.RAINFALL'])
zm_ob_test = as.vector(test_data['MEAN.ANNUAL.RAINFALL'])
zm_ob_train
```

```
$MEAN.ANNUAL.RAINFALL
  [1]  759  681  543  478  803  625  578  587  592  522  549  778  721  798  656
 [16]  580 1236  736  758  697 1626  678  741  571  810  760  688  745  914 1774
 [31] 1126  778  642  912  971  764  647  559  505  681  608  559  581  505 1444
 [46]  452  454  756  752  726  694  626  718 1046  780  580  721  700 1422  681
 [61]  794  733  729  700  956  857  836  853  778  682  606  751  744  644  656
 [76]  539 1235  505  464  853 1109  660  865  941  895  740  500 1166  641  720
 [91]  841  627  876  623  826  446  928  611  587  479  947  619  992  789  844
[106]  585  893 1294  606  550  941  700  658  647  705  629  727  637 1333  582
[121] 1092  550  663  561  553  594  732  514  601  786  657  567
```

```
zm_ob_test
```

```
$MEAN.ANNUAL.RAINFALL
 [1]  820  531  785  638  685 1061  578  520  659  735  622  674  655  695  564
[16]  648  739  533  531 1575 1316  530  612  624  612 1301  649  590  795  700
[31]  788  784  812  749  685 1238  797  551  511  556  912  698  620  559  660
[46]  974  543  955  533  605 1740  805  729  582  595  694  592  752  799  808
[61]  772 1058  535  955  741  547  684  594  572  592  543 1000
```

```
matrix_test = data.matrix(test_data[colnames(test_data) != 'MEAN.ANNUAL.RAINFALL'])
matrix_train = data.matrix(train_data[colnames(train_data) != 'MEAN.ANNUAL.RAINFALL'])
```

Task 8

```
library(GGally)
```

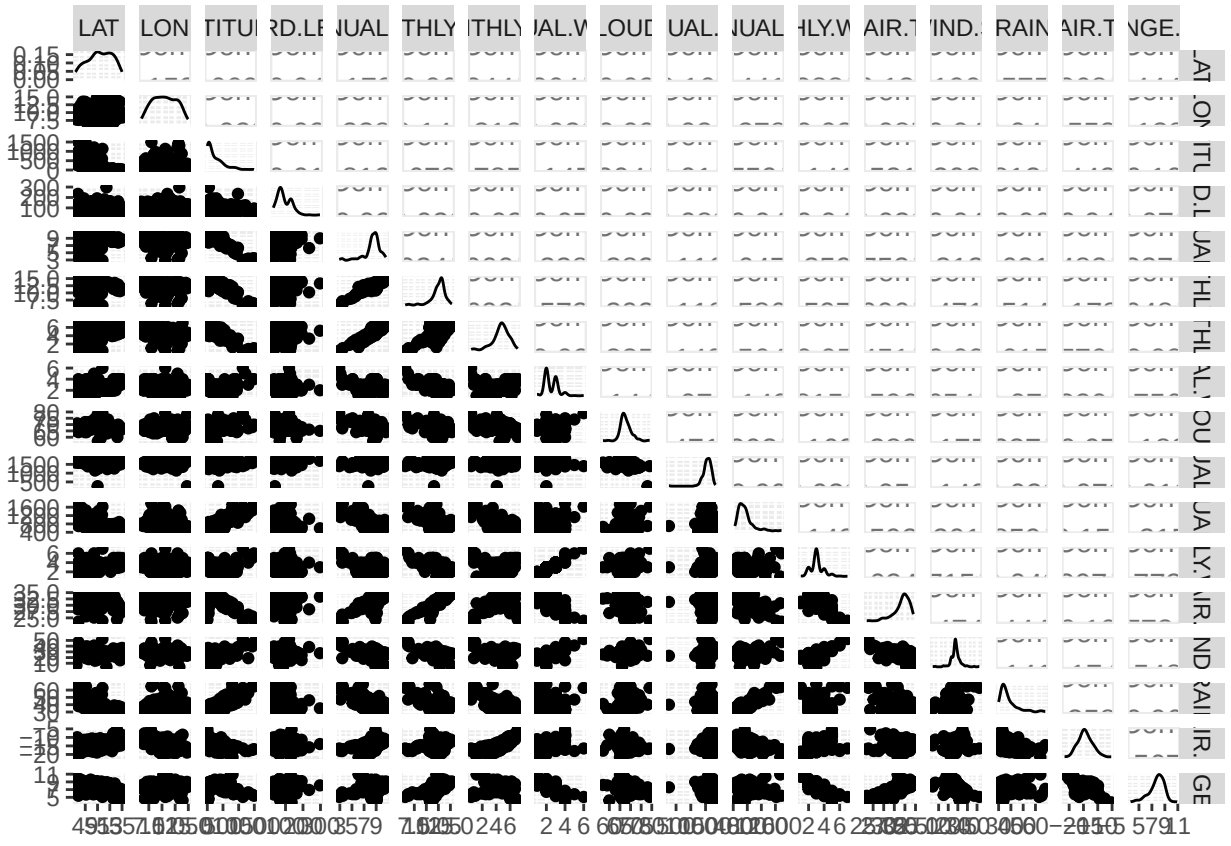
```
Warning: pakiet 'GGally' został zbudowany w wersji R 4.3.3

Ładowanie wymaganego pakietu: ggplot2

Warning: pakiet 'ggplot2' został zbudowany w wersji R 4.3.2

Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```

```
ggpairs(dane)
```



Task 9

```
correlation_matrix <- cor(dane)
correlation_matrix
```

```
                                LAT          LON     ALTITUDE RECORD.LENGTH
LAT                      1.00000000  0.15297903 -0.66889214   -0.04233468
LON                      0.15297903  1.00000000  0.02269196   -0.02188330
ALTITUDE                -0.66889214  0.02269196  1.00000000   -0.01099612
RECORD.LENGTH           -0.04233468 -0.02188330 -0.01099612    1.00000000
MEAN.ANNUAL.AIR.TEMP     0.17158381 -0.23817017 -0.81237051   -0.02505621
MEAN.MONTHLY.MAX.TEMP   -0.06008898 -0.14198037 -0.67755676    0.01996764
MEAN.MONTHLY.MIN.TEMP    0.34427310 -0.31329675 -0.78527646   -0.03888702
MEAN.ANNUAL.WIND.SPEED   0.36359362  0.01956657  0.14477716   -0.07776976
MEAN.CLOUD.COVER        -0.03225031 -0.09325165  0.30441005   -0.03518512
MEAN.ANNUAL.SUNSHINE    -0.16249067  0.03050442  0.01134756    0.08250472
MEAN.ANNUAL.RAINFALL    -0.44139472 -0.27828048  0.75853224   -0.04395966
MAX.MONTHLY.WIND.SPEED   0.38796670 -0.03281932  0.14114628   -0.04051494
MAX.AIR.TEMP            -0.13228771  0.08473182 -0.53111335    0.02980233
MAX.WIND.SPEED           0.19029701 -0.04493036  0.20802439   -0.04015454
MAX.RAINFALL            -0.57549540  0.04382930  0.81257854   -0.06283150
MIN.AIR.TEMP             0.36832117 -0.55303424 -0.44337149   -0.04521694
MEAN.RANGE.AIR.TEMP     -0.44400994  0.13762158 -0.10858197    0.07606913
```

	MEAN.ANNUAL.AIR.TEMP	MEAN.MONTHLY.MAX.TEMP
LAT	0.17158381	-0.06008898
LON	-0.23817017	-0.14198037
ALTITUDE	-0.81237051	-0.67755676
RECORD.LENGTH	-0.02505621	0.01996764
MEAN.ANNUAL.AIR.TEMP	1.00000000	0.92396613
MEAN.MONTHLY.MAX.TEMP	0.92396613	1.00000000
MEAN.MONTHLY.MIN.TEMP	0.89993502	0.69160248
MEAN.ANNUAL.WIND.SPEED	-0.33800194	-0.57238004
MEAN.CLOUD.COVER	-0.38272711	-0.39318265
MEAN.ANNUAL.SUNSHINE	0.11618803	0.11769758
MEAN.ANNUAL.RAINFALL	-0.64700505	-0.58588722
MAX.MONTHLY.WIND.SPEED	-0.35857113	-0.59708858
MAX.AIR.TEMP	0.75222067	0.90640623
MAX.WIND.SPEED	-0.31344382	-0.47103629
MAX.RAINFALL	-0.63140357	-0.51400356
MIN.AIR.TEMP	0.48164888	0.17796694
MEAN.RANGE.AIR.TEMP	0.32728804	0.64839433

	MEAN.MONTHLY.MIN.TEMP	MEAN.ANNUAL.WIND.SPEED
LAT	0.34427310	0.36359362
LON	-0.31329675	0.01956657
ALTITUDE	-0.78527646	0.14477716
RECORD.LENGTH	-0.03888702	-0.07776976
MEAN.ANNUAL.AIR.TEMP	0.89993502	-0.33800194
MEAN.MONTHLY.MAX.TEMP	0.69160248	-0.57238004
MEAN.MONTHLY.MIN.TEMP	1.00000000	-0.03254838
MEAN.ANNUAL.WIND.SPEED	-0.03254838	1.00000000
MEAN.CLOUD.COVER	-0.33482503	0.14435790
MEAN.ANNUAL.SUNSHINE	0.14012588	0.07352085
MEAN.ANNUAL.RAINFALL	-0.56402657	0.14866622
MAX.MONTHLY.WIND.SPEED	-0.05058109	0.81538673
MAX.AIR.TEMP	0.45145817	-0.59575753
MAX.WIND.SPEED	-0.09747748	0.65421507
MAX.RAINFALL	-0.61481817	0.06978958
MIN.AIR.TEMP	0.77031254	0.31972552
MEAN.RANGE.AIR.TEMP	-0.09812741	-0.75600000

	MEAN.CLOUD.COVER	MEAN.ANNUAL.SUNSHINE
LAT	-0.03225031	-0.162490674
LON	-0.09325165	0.030504424
ALTITUDE	0.30441005	0.011347564
RECORD.LENGTH	-0.03518512	0.082504718
MEAN.ANNUAL.AIR.TEMP	-0.38272711	0.116188034
MEAN.MONTHLY.MAX.TEMP	-0.39318265	0.117697577
MEAN.MONTHLY.MIN.TEMP	-0.33482503	0.140125885
MEAN.ANNUAL.WIND.SPEED	0.14435790	0.073520854
MEAN.CLOUD.COVER	1.00000000	-0.470768072
MEAN.ANNUAL.SUNSHINE	-0.47076807	1.000000000
MEAN.ANNUAL.RAINFALL	0.30224263	-0.006196494
MAX.MONTHLY.WIND.SPEED	0.16213132	0.032903035
MAX.AIR.TEMP	-0.28883397	0.054204793
MAX.WIND.SPEED	0.17451926	0.148644009
MAX.RAINFALL	0.23675529	0.029405139
MIN.AIR.TEMP	-0.07366530	0.072431802
MEAN.RANGE.AIR.TEMP	-0.18829776	0.012359004

	MEAN.ANNUAL.RAINFALL	MAX.MONTHLY.WIND.SPEED	MAX.AIR.TEMP
LAT	-0.441394718	0.38796670	-0.13228771
LON	-0.278280480	-0.03281932	0.08473182
ALTITUDE	0.758532239	0.14114628	-0.53111335
RECORD.LENGTH	-0.043959657	-0.04051494	0.02980233
MEAN.ANNUAL.AIR.TEMP	-0.647005045	-0.35857113	0.75222067
MEAN.MONTHLY.MAX.TEMP	-0.585887219	-0.59708858	0.90640623
MEAN.MONTHLY.MIN.TEMP	-0.564026570	-0.05058109	0.45145817
MEAN.ANNUAL.WIND.SPEED	0.148666215	0.81538673	-0.59575753
MEAN.CLOUD.COVER	0.302242631	0.16213132	-0.28883397
MEAN.ANNUAL.SUNSHINE	-0.006196494	0.03290304	0.05420479
MEAN.ANNUAL.RAINFALL	1.000000000	0.14253058	-0.59173837
MAX.MONTHLY.WIND.SPEED	0.142530583	1.00000000	-0.62414738
MAX.AIR.TEMP	-0.591738374	-0.62414738	1.00000000
MAX.WIND.SPEED	0.200972473	0.71527966	-0.47446238
MAX.RAINFALL	0.858771353	0.04911154	-0.44440093
MIN.AIR.TEMP	-0.158892303	0.30718138	-0.10559467
MEAN.RANGE.AIR.TEMP	-0.215098374	-0.77260900	0.77335956

	MAX.WIND.SPEED	MAX.RAINFALL	MIN.AIR.TEMP
LAT	0.19029701	-0.57549540	0.36832117
LON	-0.04493036	0.04382930	-0.55303424
ALTITUDE	0.20802439	0.81257854	-0.44337149
RECORD.LENGTH	-0.04015454	-0.06283150	-0.04521694
MEAN.ANNUAL.AIR.TEMP	-0.31344382	-0.63140357	0.48164888
MEAN.MONTHLY.MAX.TEMP	-0.47103629	-0.51400356	0.17796694
MEAN.MONTHLY.MIN.TEMP	-0.09747748	-0.61481817	0.77031254
MEAN.ANNUAL.WIND.SPEED	0.65421507	0.06978958	0.31972552
MEAN.CLOUD.COVER	0.17451926	0.23675529	-0.07366530
MEAN.ANNUAL.SUNSHINE	0.14864401	0.02940514	0.07243180
MEAN.ANNUAL.RAINFALL	0.20097247	0.85877135	-0.15889230
MAX.MONTHLY.WIND.SPEED	0.71527966	0.04911154	0.30718138
MAX.AIR.TEMP	-0.47446238	-0.44440093	-0.10559467
MAX.WIND.SPEED	1.00000000	0.14068137	0.17421081
MAX.RAINFALL	0.14068137	1.00000000	-0.35782557
MIN.AIR.TEMP	0.17421081	-0.35782557	1.00000000
MEAN.RANGE.AIR.TEMP	-0.54822300	-0.06189922	-0.56682858

	MEAN.RANGE.AIR.TEMP
LAT	-0.44400994
LON	0.13762158
ALTITUDE	-0.10858197
RECORD.LENGTH	0.07606913
MEAN.ANNUAL.AIR.TEMP	0.32728804
MEAN.MONTHLY.MAX.TEMP	0.64839433
MEAN.MONTHLY.MIN.TEMP	-0.09812741
MEAN.ANNUAL.WIND.SPEED	-0.75600000
MEAN.CLOUD.COVER	-0.18829776
MEAN.ANNUAL.SUNSHINE	0.01235900
MEAN.ANNUAL.RAINFALL	-0.21509837
MAX.MONTHLY.WIND.SPEED	-0.77260900
MAX.AIR.TEMP	0.77335956
MAX.WIND.SPEED	-0.54822300
MAX.RAINFALL	-0.06189922
MIN.AIR.TEMP	-0.56682858
MEAN.RANGE.AIR.TEMP	1.00000000

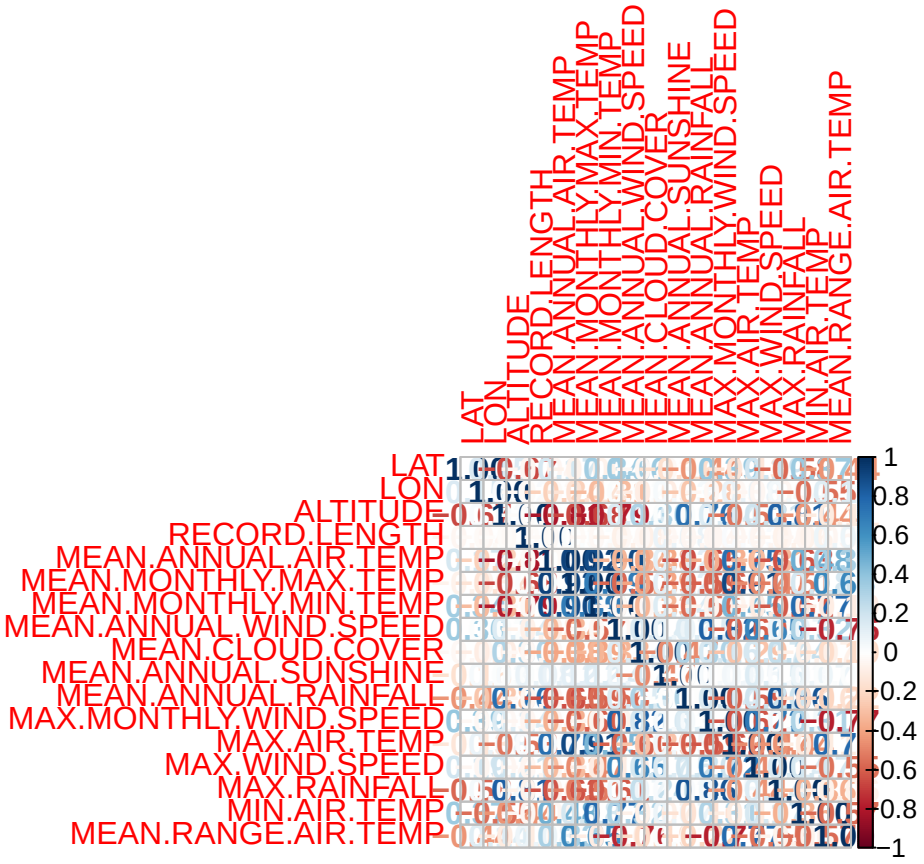
Task 10

```
library(corrplot)
```

```
  Warning: pakiet 'corrplot' został zbudowany w wersji R 4.3.2
```

```
  corrplot 0.92 loaded
```

```
corrplot(correlation_matrix, method = "number")
```



Task 11

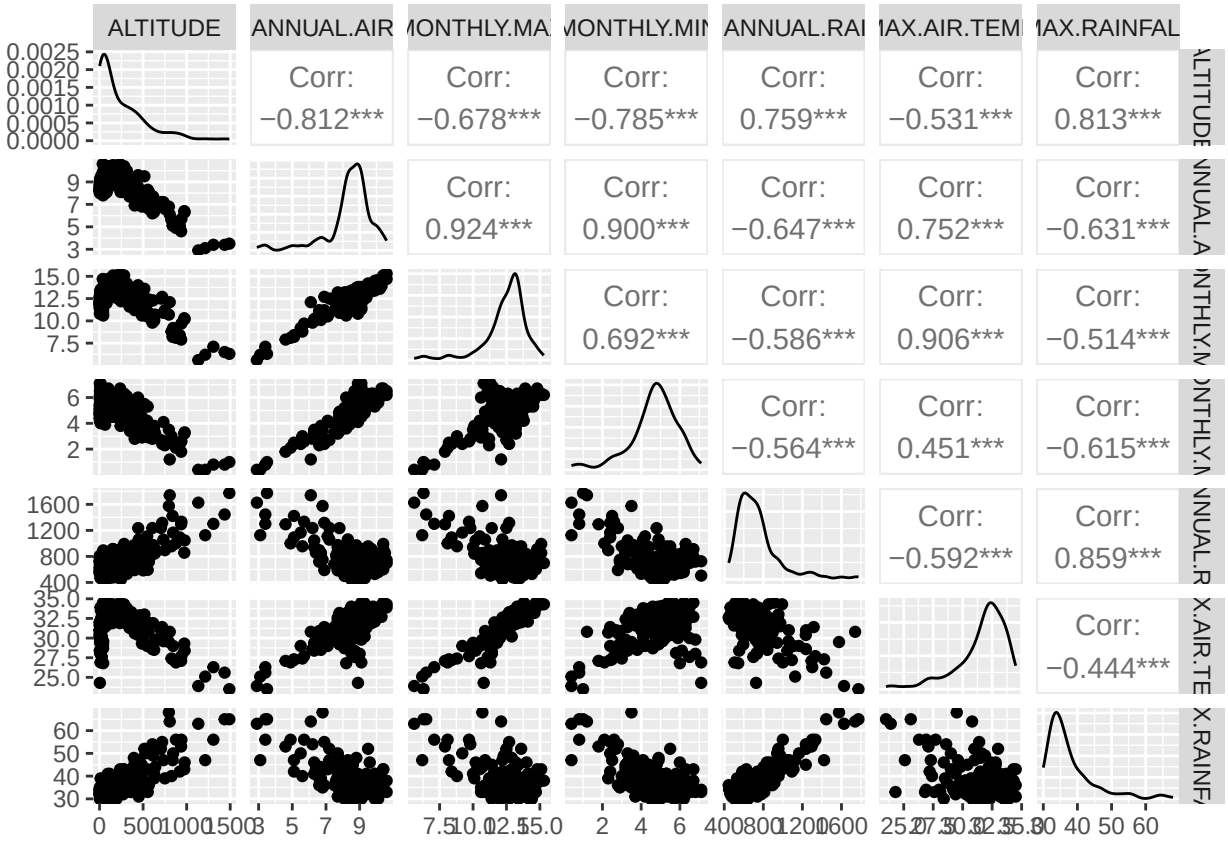
```
filtr <- abs(cor(dane['MEAN.ANNUAL.RAINFALL'],dane)) >= 0.5
filtr
```

```
                          LAT   LON ALTITUDE RECORD.LENGTH MEAN.ANNUAL.AIR.TEMP
  MEAN.ANNUAL.RAINFALL FALSE FALSE     TRUE          FALSE                 TRUE
                       MEAN.MONTHLY.MAX.TEMP MEAN.MONTHLY.MIN.TEMP
  MEAN.ANNUAL.RAINFALL                  TRUE                  TRUE
                       MEAN.ANNUAL.WIND.SPEED MEAN.CLOUD.COVER
  MEAN.ANNUAL.RAINFALL                  FALSE            FALSE
```

```
                      MEAN.ANNUAL.SUNSHINE MEAN.ANNUAL.RAINFALL
MEAN.ANNUAL.RAINFALL                FALSE                  TRUE
                      MAX.MONTHLY.WIND.SPEED MAX.AIR.TEMP MAX.WIND.SPEED
MEAN.ANNUAL.RAINFALL                  FALSE         TRUE          FALSE
                      MAX.RAINFALL MIN.AIR.TEMP MEAN.RANGE.AIR.TEMP
MEAN.ANNUAL.RAINFALL         TRUE        FALSE                FALSE
```

Task 12

```
library(GGally)
ggpairs(dane[as.vector(filtr)])
```



Part 2

Task 1

```
model = lm(MEAN.ANNUAL.RAINFALL ~ 1, train_data)[1]
model<- as.double(model)
```

```r
srednia<-mean(dane$MEAN.ANNUAL.RAINFALL)
wyniki<-c(srednia,model)
wyniki
```

```
[1] 745.6569 749.3788
```

Task 2

```r
rmse_df <- data.frame(
  Model = character(),
  RMSE_Train = numeric(),
  RMSE_Test = numeric(),
  stringsAsFactors = FALSE
)
```

Task 3

```r
library(Metrics)
```

```
Warning: pakiet 'Metrics' został zbudowany w wersji R 4.3.3
```

```r
predicted<-lm(MEAN.ANNUAL.RAINFALL ~ 1, train_data)
observed<-predicted$fitted.values
predicted <- train_data$MEAN.ANNUAL.RAINFALL
rmse_train1<- sqrt(mean((predicted-observed)^2))
rmse_train1
```

```
[1] 228.8048
```

Task 4

```r
model = lm(MEAN.ANNUAL.RAINFALL ~ 1, train_data)
predicted = predict(model, test_data)
observed = test_data$MEAN.ANNUAL.RAINFALL
rmse_test1<-sqrt(mean((predicted - observed)^2))
rmse_test1
```

```
[1] 236.9799
```

Task 5

```r
model2 = lm(MEAN.ANNUAL.RAINFALL ~ ALTITUDE, train_data)
predicted2 = model2[5]
observed2 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train2<-sqrt(mean((predicted2$fitted.values - observed2)^2))
```

Task 6

```
predicted2 = predict(model2, test_data)
observed2 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test2<-sqrt(mean((predicted2 - observed2)^2))
rmse_test2
```

```
[1] 163.6962
```

Task 7

```
model3 = lm(MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL, train_data)
predicted3 = model3[5]
observed3 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train3<-sqrt(mean((predicted3$fitted.values - observed3)^2))
rmse_train3
```

```
[1] 120.9056
```

Task 8

```
predicted3 = predict(model3, test_data)
observed3 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test3<-sqrt(mean((predicted3 - observed3)^2))
rmse_test3
```

```
[1] 115.2687
```

Task 9

```
model4 = lm(MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + ALTITUDE , train_data)
predicted4 = model4[5]
observed4 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train4<-sqrt(mean((predicted4$fitted.values - observed4)^2))
rmse_train4
```

```
[1] 117.3791
```

```
r_squared <- summary(model4)$r.squared
r_squared
```

```
[1] 0.7368209
```

```
#R² is 0.7289122, which means that approximately 72.89% of the variability in the annual average precip
```

Task 10

```
predicted4 = predict(model4, test_data)
observed4 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test4<-sqrt(mean((predicted4 - observed4)^2))
rmse_test4
```

```
[1] 114.451
```

Task 11

```
rmse_df <- data.frame(
  Model = c("Bazowy", "Regresja liniowa (ALTITUDE)", "Regresja liniowa (MAX.RAINFALL)", "Regresja liniow
  RMSE_Train = c(rmse_train1, rmse_train2,rmse_train3,rmse_train4),
  RMSE_Test = c(rmse_test1, rmse_test2,rmse_test3,rmse_test4)
)

print(rmse_df)
```

```
                          Model RMSE_Train RMSE_Test
1                        Bazowy   228.8048  236.9799
2     Regresja liniowa (ALTITUDE)   143.8115  163.6962
3 Regresja liniowa (MAX.RAINFALL)   120.9056  115.2687
4    Regresja liniowa wielokrotna   117.3791  114.4510
```

Part 3

Task 1

```
model5 = lm(MEAN.ANNUAL.RAINFALL ~ ALTITUDE + log(ALTITUDE) + ALTITUDE*ALTITUDE, train_data)
predicted5 = model5[5]
observed5 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train5<-sqrt(mean((predicted5$fitted.values - observed5)^2))
rmse_train5
```

```
[1] 135.0549
```

Task 2

```
predicted5 = predict(model5, test_data)
observed5 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test5<-sqrt(mean((predicted5 - observed5)^2))
rmse_test5
```

```
[1] 153.4564
```

Task 3

```
model6 = lm(MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + log(MAX.RAINFALL) + MAX.RAINFALL*MAX.RAINFALL, train_
predicted6 = model6[5]
observed6 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train5<-sqrt(mean((predicted6$fitted.values - observed6)^2))
rmse_train5
```

```
[1] 117.5012
```

Task 4

```
predicted6 = predict(model6, test_data)
observed6 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test6<-sqrt(mean((predicted6 - observed6)^2))
rmse_test6
```

```
[1] 111.6988
```

Task 5

```
model7 = lm(MEAN.ANNUAL.RAINFALL ~ 1, train_data)
step(model7, scope = as.formula(MEAN.ANNUAL.RAINFALL ~ ALTITUDE + MAX.RAINFALL + MEAN.CLOUD.COVER + MEA
```

```
Start:  AIC=1436.28
MEAN.ANNUAL.RAINFALL ~ 1

                       Df Sum of Sq     RSS    AIC
+ MAX.RAINFALL          1   4980814 1929599 1269.9
+ ALTITUDE              1   4180423 2729990 1315.7
+ MEAN.ANNUAL.AIR.TEMP  1   3287145 3623268 1353.0
+ MEAN.CLOUD.COVER      1   1329273 5581140 1410.1
<none>                              6910413 1436.3

Step:  AIC=1269.88
MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL

                       Df Sum of Sq     RSS    AIC
```

```
+ MEAN.ANNUAL.AIR.TEMP  1    211879 1717720 1256.5
+ MEAN.CLOUD.COVER      1    155870 1773729 1260.8
+ ALTITUDE              1    110922 1818677 1264.1
<none>                              1929599 1269.9

Step:  AIC=1256.53
MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.ANNUAL.AIR.TEMP

                   Df Sum of Sq     RSS    AIC
+ MEAN.CLOUD.COVER  1     65671 1652049 1253.4
<none>                          1717720 1256.5
+ ALTITUDE          1       448 1717272 1258.5

Step:  AIC=1253.38
MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.ANNUAL.AIR.TEMP +
    MEAN.CLOUD.COVER

           Df Sum of Sq     RSS    AIC
<none>                  1652049 1253.4
+ ALTITUDE  1    1367.1 1650682 1255.3


Call:
lm(formula = MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.ANNUAL.AIR.TEMP +
    MEAN.CLOUD.COVER, data = train_data)

Coefficients:
         (Intercept)          MAX.RAINFALL  MEAN.ANNUAL.AIR.TEMP
            -483.378                23.539               -29.162
    MEAN.CLOUD.COVER
               8.739
```

Task 6

```
model8 = lm(MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.ANNUAL.AIR.TEMP +
MEAN.CLOUD.COVER, data = train_data)
predicted8 = model8[5]
observed8 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train8<-sqrt(mean((predicted8$fitted.values - observed8)^2))
rmse_train8
```

```
[1] 111.8728
```

Task 7

```
predicted8 = predict(model8, test_data)
observed8 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test8<-sqrt(mean((predicted8 - observed8)^2))
rmse_test8
```

```
[1] 117.8248
```

Task 8

```
model9 = lm(MEAN.ANNUAL.RAINFALL ~ ALTITUDE + MAX.RAINFALL + MEAN.CLOUD.COVER + MEAN.ANNUAL.AIR.TEMP, t
step(model9, direction = 'backward')
```

```
Start:  AIC=1255.27
MEAN.ANNUAL.RAINFALL ~ ALTITUDE + MAX.RAINFALL + MEAN.CLOUD.COVER +
    MEAN.ANNUAL.AIR.TEMP

                       Df Sum of Sq     RSS    AIC
- ALTITUDE              1      1367 1652049 1253.4
<none>                              1650682 1255.3
- MEAN.ANNUAL.AIR.TEMP  1     53555 1704237 1257.5
- MEAN.CLOUD.COVER      1     66590 1717272 1258.5
- MAX.RAINFALL          1    921839 2572521 1311.8

Step:  AIC=1253.38
MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.CLOUD.COVER + MEAN.ANNUAL.AIR.TEMP

                       Df Sum of Sq     RSS    AIC
<none>                              1652049 1253.4
- MEAN.CLOUD.COVER      1     65671 1717720 1256.5
- MEAN.ANNUAL.AIR.TEMP  1    121679 1773729 1260.8
- MAX.RAINFALL          1   1849254 3501303 1350.5


Call:
lm(formula = MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.CLOUD.COVER +
    MEAN.ANNUAL.AIR.TEMP, data = train_data)

Coefficients:
         (Intercept)          MAX.RAINFALL      MEAN.CLOUD.COVER
            -483.378                23.539                 8.739
MEAN.ANNUAL.AIR.TEMP
             -29.162
```

Task 9

```
model10 = lm(MEAN.ANNUAL.RAINFALL ~ MAX.RAINFALL + MEAN.ANNUAL.AIR.TEMP +
MEAN.CLOUD.COVER, data = train_data)
predicted10 = model10[5]
observed10 = train_data$MEAN.ANNUAL.RAINFALL
rmse_train10<-sqrt(mean((predicted10$fitted.values - observed10)^2))
rmse_train10
```

```
[1] 111.8728
```

Task 10

```
predicted10 = predict(model10, test_data)
observed10 = test_data$MEAN.ANNUAL.RAINFALL
rmse_test10<-sqrt(mean((predicted10 - observed10)^2))
rmse_test10
```

```
[1] 117.8248
```