

Exploratory analysis

```
# getwd()
# setwd('DATA 583/583-project')

# Load the data from the CSV file
data <- read.csv("data_whole.csv", header = TRUE, check.names = FALSE)
dim(data) # 2857, 33
```

```
## [1] 2857 33
```

```
## data cleaning
# remove y2, x6, x8, x9, x22, x29 because of missing values for some rows
data <- data[, -c(3, 9, 11, 12, 25, 32)]

# remove rows with missing values, obtain only weekly data, rather than daily
data <- na.omit(data)
dim(data) # 352, 27
```

```
## [1] 352 27
```

```
# rename y1 to y
names(data)[2] <- "y"

head(data)
```

```
##      date    y    x1    x2  x3      x4    x5      x7    x10
## 1079 2016/1/15 4380 29.42 28.94 370  -47.12009 -14.08 -1313.2016 312.1795
## 1084 2016/1/22 4500 32.19 32.18 370  -83.91724 -30.60 -1019.3778 312.1795
## 1089 2016/1/29 4715 33.62 34.74 375 -226.87681 -43.40  -637.3231 397.6496
## 1094 2016/2/5  4710 30.89 34.06 380 -214.94086 -46.42  -470.3440 397.6496
## 1099 2016/2/12 4710 29.44 33.36 380 -180.44046 -42.46  -466.3399 372.6496
## 1104 2016/2/19 5075 29.64 33.01 380  -72.41137 -34.74  -261.4895 408.1197
##      x11    x12    x13    x14 x15    x16  x17  x18  x19 x20 x21  x23  x24
## 1079 -173.8   26.2  186.2  211.2  90   111.2  70.7  66.4  18  57  49  70.3  15.3
## 1084 -274.8  -34.8   60.2   95.2  75    10.2  68.4  65.9  18  48  38  64.6  15.6
## 1089 -256.5   -6.5   -1.5   13.5  55    33.5  65.7  60.3  15   9   7  60.0  14.1
## 1094 -246.2    3.8    8.8   23.8  55    43.8  65.7  60.3  15   9   7  60.0  14.1
## 1099 -246.2    3.8    8.8   23.8  55    43.8  62.5  60.3  15   9   7  60.0  14.1
## 1104 -377.8 -127.8 -142.8 -132.8 95  -127.8  63.5  59.8  15  23  37  53.9  18.4
##      x25  x26  x27  x28    x30
## 1079 21.0  7.2  60.1  7.0  8.72712
## 1084 21.7  7.8  60.6  7.7  8.90368
## 1089 21.8  7.8  51.6  7.6  9.24000
## 1094 21.8  7.8  51.6  6.9  8.91000
## 1099 21.8  7.8  51.6  7.3  7.68750
## 1104 30.0 16.5  79.7  7.4  6.90140
```

```
# The structure of the data,
# including the number of variables, their types, and the first few observations.
str(data)
```

```
## 'data.frame': 352 obs. of 27 variables:
## $ date: chr "2016/1/15" "2016/1/22" "2016/1/29" "2016/2/5" ...
## $ y : num 4380 4500 4715 4710 4710 ...
## $ x1 : num 29.4 32.2 33.6 30.9 29.4 ...
## $ x2 : num 28.9 32.2 34.7 34.1 33.4 ...
## $ x3 : num 370 370 375 380 380 380 380 385 390 390 ...
## $ x4 : num -47.1 -83.9 -226.9 -214.9 -180.4 ...
## $ x5 : num -14.1 -30.6 -43.4 -46.4 -42.5 ...
## $ x7 : num -1313 -1019 -637 -470 -466 ...
## $ x10 : num 312 312 398 398 373 ...
## $ x11 : num -174 -275 -256 -246 -246 ...
## $ x12 : num 26.2 -34.8 -6.5 3.8 3.8 ...
## $ x13 : num 186.2 60.2 -1.5 8.8 8.8 ...
## $ x14 : num 211.2 95.2 13.5 23.8 23.8 ...
## $ x15 : int 90 75 55 55 55 95 80 255 135 160 ...
## $ x16 : num 111.2 10.2 33.5 43.8 43.8 ...
## $ x17 : num 70.7 68.4 65.7 65.7 62.5 63.5 69.9 74.3 78 80.7 ...
## $ x18 : num 66.4 65.9 60.3 60.3 60.3 59.8 67.6 74 78.9 80.7 ...
## $ x19 : num 18 18 15 15 15 15 20 30 36 39 ...
## $ x20 : num 57 48 9 9 9 23 61 76 82 82 ...
## $ x21 : num 49 38 7 7 7 37 63 74 79 79 ...
## $ x23 : num 70.3 64.6 60 60 60 53.9 60.8 65.4 69.2 69 ...
## $ x24 : num 15.3 15.6 14.1 14.1 14.1 18.4 18.1 17 13.2 13.7 ...
## $ x25 : num 21 21.7 21.8 21.8 21.8 30 29.6 28.6 25.2 25 ...
## $ x26 : num 7.2 7.8 7.8 7.8 7.8 16.5 17.5 15.5 10.1 9.9 ...
## $ x27 : num 60.1 60.6 51.6 51.6 51.6 79.7 79.8 69.7 53.5 54.5 ...
## $ x28 : num 7 7.7 7.6 6.9 7.3 7.4 9 8.3 7.7 11.1 ...
## $ x30 : num 8.73 8.9 9.24 8.91 7.69 ...
## - attr(*, "na.action")= 'omit' Named int [1:2505] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:2505] "1" "2" "3" "4" ...
```

```
summary(data)
```

```
##      date              y              x1              x2
## Length:352      Min.   :3130      Min.   : 16.94      Min.   : 21.44
## Class :character 1st Qu.:4428      1st Qu.: 47.73      1st Qu.: 49.97
## Mode  :character Median :5118      Median : 56.70      Median : 63.34
##              Mean   :5427      Mean   : 60.18      Mean   : 64.51
##              3rd Qu.:6511      3rd Qu.: 69.64      3rd Qu.: 75.33
##              Max.   :8440      Max.   :121.37      Max.   :122.95
##      x3              x4              x5              x7
## Min.   : 370.0      Min.   :-688.89      Min.   :-56.480      Min.   : -1742.1
## 1st Qu.: 575.0      1st Qu.: -151.02      1st Qu.: -17.983      1st Qu.: -908.7
## Median : 604.8      Median : -17.50      Median :  0.600      Median : -348.5
## Mean   : 747.8      Mean   :  58.92      Mean   :  7.387      Mean   : -287.3
## 3rd Qu.: 900.0      3rd Qu.: 174.37      3rd Qu.: 22.125      3rd Qu.:  244.5
## Max.   :2350.0      Max.   :1754.92      Max.   :128.740      Max.   : 1658.5
##      x10              x11              x12              x13
## Min.   : -372.6      Min.   : -420.80      Min.   : -585.23      Min.   : -745.23
```

```
## 1st Qu.: 204.4 1st Qu.: -73.16 1st Qu.: 30.88 1st Qu.: -29.79
## Median : 767.6 Median : 26.89 Median : 170.20 Median : 168.22
## Mean : 861.3 Mean : 43.85 Mean : 315.46 Mean : 201.25
## 3rd Qu.:1166.6 3rd Qu.: 145.03 3rd Qu.: 472.56 3rd Qu.: 430.65
## Max. :3516.2 Max. : 573.15 Max. :2370.97 Max. :1188.15
## x14 x15 x16 x17
## Min. :-715.2 Min. :-180.0 Min. :-234.5 Min. :59.50
## 1st Qu.: -114.4 1st Qu.: 145.0 1st Qu.: 107.7 1st Qu.:82.90
## Median : 138.7 Median : 270.0 Median : 270.2 Median :87.60
## Mean : 163.6 Mean : 282.9 Mean : 316.9 Mean :86.13
## 3rd Qu.: 418.9 3rd Qu.: 410.0 3rd Qu.: 507.5 3rd Qu.:91.20
## Max. :1246.7 Max. : 920.0 Max. :1313.1 Max. :97.70
## x18 x19 x20 x21
## Min. :47.50 Min. :13.30 Min. : 4.00 Min. : 1.00
## 1st Qu.:78.50 1st Qu.:41.00 1st Qu.:68.00 1st Qu.:59.00
## Median :85.00 Median :47.00 Median :78.00 Median :70.00
## Mean :82.72 Mean :47.27 Mean :72.82 Mean :65.35
## 3rd Qu.:89.12 3rd Qu.:54.92 3rd Qu.:85.00 3rd Qu.:78.00
## Max. :97.10 Max. :84.00 Max. :95.00 Max. :93.00
## x23 x24 x25 x26
## Min. : 42.50 Min. : 2.00 Min. : 7.30 Min. : 0.80
## 1st Qu.: 78.90 1st Qu.:11.65 1st Qu.:16.68 1st Qu.: 7.50
## Median : 83.60 Median :16.10 Median :21.70 Median :11.80
## Mean : 83.16 Mean :16.79 Mean :21.99 Mean :13.07
## 3rd Qu.: 89.05 3rd Qu.:22.23 3rd Qu.:26.90 3rd Qu.:16.73
## Max. :100.40 Max. :31.10 Max. :38.60 Max. :33.40
## x27 x28 x30
## Min. : 10.70 Min. : 2.500 Min. : 6.77
## 1st Qu.: 40.75 1st Qu.: 6.100 1st Qu.:10.39
## Median : 54.75 Median : 7.700 Median :11.92
## Mean : 58.56 Mean : 8.263 Mean :11.97
## 3rd Qu.: 70.92 3rd Qu.:10.000 3rd Qu.:13.42
## Max. :114.00 Max. :19.000 Max. :17.38
```

```
# Create a correlation matrix for the explanatory variables
cor(data[,2:27], use = "pairwise.complete.obs")
```

```
## y x1 x2 x3 x4 x5
## y 1.00000000 0.052031837 0.06362446 -0.19147023 0.21665727 0.24219174
## x1 0.05203184 1.000000000 0.99177105 0.76609575 -0.09154875 0.35389070
## x2 0.06362446 0.991771045 1.00000000 0.74457003 -0.02984453 0.42277449
## x3 -0.19147023 0.766095749 0.74457003 1.00000000 -0.20559418 0.10243817
## x4 0.21665727 -0.091548748 -0.02984453 -0.20559418 1.00000000 0.68375110
## x5 0.24219174 0.353890695 0.42277449 0.10243817 0.68375110 1.00000000
## x7 0.61729838 -0.112705950 -0.08457051 -0.15839040 0.23860317 0.11215846
## x10 0.53046401 0.180686960 0.14733905 0.12775041 -0.04658959 -0.09659268
## x11 0.05922413 -0.053187517 -0.01160951 -0.11258935 0.03295098 0.16208444
## x12 -0.04981699 0.335303064 0.34658502 0.19486212 -0.10470183 0.30704024
## x13 0.38547638 -0.060822189 -0.04275167 -0.18826174 0.14607847 0.23117081
## x14 0.38860790 -0.203443629 -0.18270622 -0.30592352 -0.01240553 0.07663304
## x15 -0.07353146 -0.209223452 -0.22142431 -0.00793429 -0.31909119 -0.23238002
## x16 0.14735097 -0.303744152 -0.25254556 -0.29471411 0.11062361 0.14306130
## x17 0.19742925 0.007687296 0.02548198 -0.12053847 0.31539564 0.33495935
## x18 0.50128494 -0.199801855 -0.18270143 -0.37735746 0.36457807 0.27126478
```

```

## x19 -0.09073899 0.202451651 0.22394658 0.13772882 0.19017367 0.35162662
## x20 0.06001015 -0.102753414 -0.10965315 -0.10516177 0.09122113 0.08461802
## x21 0.08853691 -0.152658810 -0.15629960 -0.15701705 0.11654304 0.10038800
## x23 -0.11406186 -0.205945242 -0.20804401 -0.13289299 0.06666770 -0.01848611
## x24 -0.46648209 0.517315396 0.49320412 0.63068356 -0.35432260 0.00045600
## x25 -0.30023629 0.316450494 0.30044503 0.27578071 -0.27204492 0.02270305
## x26 -0.53337932 0.477301556 0.45954961 0.56580105 -0.35377688 -0.03608236
## x27 -0.39613873 0.478141186 0.46116589 0.51062091 -0.32098581 0.04868221
## x28 -0.37819640 -0.028053955 -0.05956647 -0.03796728 -0.13697554 -0.15929351
## x30 -0.52986062 0.100007307 0.12030340 0.23151843 0.11303493 0.13643980
##          x7          x10          x11          x12          x13          x14
## y    0.61729838 0.5304640130 0.05922413 -0.04981699 0.38547638 0.38860790
## x1   -0.11270595 0.1806869598 -0.05318752 0.33530306 -0.06082219 -0.20344363
## x2   -0.08457051 0.1473390469 -0.01160951 0.34658502 -0.04275167 -0.18270622
## x3   -0.15839040 0.1277504081 -0.11258935 0.19486212 -0.18826174 -0.30592352
## x4    0.23860317 -0.0465895891 0.03295098 -0.10470183 0.14607847 -0.01240553
## x5    0.11215846 -0.0965926831 0.16208444 0.30704024 0.23117081 0.07663304
## x7    1.00000000 0.2241152778 0.12875656 -0.06878916 0.30294842 0.21909024
## x10   0.22411528 1.0000000000 -0.02760491 -0.25973562 0.09424150 0.05197156
## x11   0.12875656 -0.0276049141 1.00000000 0.19734205 0.48779274 0.53579558
## x12  -0.06878916 -0.2597356201 0.19734205 1.00000000 0.16658867 0.19077477
## x13   0.30294842 0.0942414969 0.48779274 0.16658867 1.00000000 0.72906536
## x14   0.21909024 0.0519715620 0.53579558 0.19077477 0.72906536 1.00000000
## x15  -0.04147202 -0.1418191916 0.01012681 0.02710399 -0.01324694 0.06506147
## x16   0.21869633 -0.1376876336 0.70309747 0.17302865 0.41126245 0.47503372
## x17   0.15793505 0.0114027244 0.06395389 0.15315452 0.25409895 0.02609498
## x18   0.31207264 0.2342165187 0.06913204 -0.06682643 0.34723241 0.18944188
## x19  -0.04900836 -0.0820677689 0.05095321 0.20486468 0.22001784 -0.04522346
## x20  -0.07959492 0.1340124286 0.02685174 -0.01619101 0.20249448 0.10111165
## x21  -0.09662947 0.0873412859 0.03014270 -0.03659966 0.27686967 0.18114431
## x23   0.03883573 -0.1650150508 0.05861706 0.13177402 0.12071823 0.03427904
## x24  -0.39629664 -0.0378581082 -0.04528022 0.23595073 -0.23352422 -0.36545971
## x25  -0.33795285 0.0092896233 -0.05099725 0.24715123 -0.32667096 -0.24255814
## x26  -0.44872736 -0.0847635398 0.03107040 0.26710755 -0.36032976 -0.34671065
## x27  -0.41120106 0.0009661917 -0.02740642 0.26519139 -0.32610398 -0.31454603
## x28  -0.34211609 -0.2136970776 -0.01880267 0.20194580 -0.24422753 -0.17992628
## x30  -0.20291040 -0.4571345814 0.10903473 0.29815346 -0.13144087 -0.31860696
##          x15          x16          x17          x18          x19          x20
## y   -0.073531461 0.147350973 0.197429249 0.50128494 -0.09073899 0.06001015
## x1   -0.209223452 -0.303744152 0.007687296 -0.19980185 0.20245165 -0.10275341
## x2   -0.221424311 -0.252545563 0.025481978 -0.18270143 0.22394658 -0.10965315
## x3   -0.007934290 -0.294714107 -0.120538466 -0.37735746 0.13772882 -0.10516177
## x4   -0.319091187 0.110623606 0.315395642 0.36457807 0.19017367 0.09122113
## x5   -0.232380017 0.143061298 0.334959348 0.27126478 0.35162662 0.08461802
## x7   -0.041472022 0.218696330 0.157935048 0.31207264 -0.04900836 -0.07959492
## x10  -0.141819192 -0.137687634 0.011402724 0.23421652 -0.08206777 0.13401243
## x11   0.010126814 0.703097465 0.063953887 0.06913204 0.05095321 0.02685174
## x12   0.027103992 0.173028653 0.153154516 -0.06682643 0.20486468 -0.01619101
## x13  -0.013246942 0.411262454 0.254098954 0.34723241 0.22001784 0.20249448
## x14   0.065061471 0.475033717 0.026094975 0.18944188 -0.04522346 0.10111165
## x15   1.000000000 0.148198317 -0.009786551 0.03600575 0.03932110 0.03295713
## x16   0.148198317 1.000000000 0.259066458 0.23170607 0.08369908 0.07864882
## x17  -0.009786551 0.259066458 1.000000000 0.85535685 0.76931579 0.70696844
## x18   0.036005755 0.231706066 0.855356852 1.00000000 0.55347492 0.66809088

```

```

## x19 0.039321105 0.083699077 0.769315793 0.55347492 1.00000000 0.76887769
## x20 0.032957135 0.078648824 0.706968439 0.66809088 0.76887769 1.00000000
## x21 0.032318067 0.108245374 0.659602239 0.64411585 0.70771583 0.94644881
## x23 0.094083532 0.291792979 0.723731061 0.46495023 0.59179557 0.55870540
## x24 0.080523451 -0.271078182 -0.190930983 -0.43898021 0.10345701 -0.09837752
## x25 -0.208026440 -0.170757582 -0.189921974 -0.35053897 -0.08800486 -0.16324502
## x26 0.015367197 -0.212383495 -0.265978341 -0.51537835 0.02456802 -0.12796179
## x27 -0.074224614 -0.263208009 -0.274753906 -0.46263252 -0.04231405 -0.16890905
## x28 -0.031216974 0.009879839 -0.055645346 -0.21161189 -0.09423986 -0.03351619
## x30 -0.025704507 0.231499103 0.291966182 -0.11257880 0.29756617 0.02455556
##      x21      x23      x24      x25      x26
## y 0.08853691 -0.1140618586 -0.46648209 -0.300236293 -0.53337932
## x1 -0.15265881 -0.2059452417 0.51731540 0.316450494 0.47730156
## x2 -0.15629960 -0.2080440147 0.49320412 0.300445031 0.45954961
## x3 -0.15701705 -0.1328929854 0.63068356 0.275780707 0.56580105
## x4 0.11654304 0.0666677010 -0.35432260 -0.272044915 -0.35377688
## x5 0.10038800 -0.0184861092 0.00045600 0.022703048 -0.03608236
## x7 -0.09662947 0.0388357270 -0.39629664 -0.337952846 -0.44872736
## x10 0.08734129 -0.1650150508 -0.03785811 0.009289623 -0.08476354
## x11 0.03014270 0.0586170607 -0.04528022 -0.050997251 0.03107040
## x12 -0.03659966 0.1317740177 0.23595073 0.247151226 0.26710755
## x13 0.27686967 0.1207182305 -0.23352422 -0.326670962 -0.36032976
## x14 0.18114431 0.0342790412 -0.36545971 -0.242558142 -0.34671065
## x15 0.03231807 0.0940835318 0.08052345 -0.208026440 0.01536720
## x16 0.10824537 0.2917929787 -0.27107818 -0.170757582 -0.21238349
## x17 0.65960224 0.7237310608 -0.19093098 -0.189921974 -0.26597834
## x18 0.64411585 0.4649502255 -0.43898021 -0.350538972 -0.51537835
## x19 0.70771583 0.5917955711 0.10345701 -0.088004863 0.02456802
## x20 0.94644881 0.5587054037 -0.09837752 -0.163245018 -0.12796179
## x21 1.00000000 0.5231570897 -0.14621266 -0.185323009 -0.15817773
## x23 0.52315709 1.0000000000 -0.19632170 -0.248283673 -0.23847943
## x24 -0.14621266 -0.1963216974 1.00000000 0.735629032 0.90451706
## x25 -0.18532301 -0.2482836725 0.73562903 1.000000000 0.78196595
## x26 -0.15817773 -0.2384794304 0.90451706 0.781965951 1.00000000
## x27 -0.19296359 -0.3261944813 0.91178030 0.907001010 0.93399664
## x28 -0.03088271 0.0004736721 0.25474175 0.416842847 0.36844012
## x30 -0.03269500 0.4109034121 0.34580884 0.246156388 0.34027428
##      x27      x28      x30
## y -0.3961387278 -0.3781964014 -0.52986062
## x1 0.4781411856 -0.0280539548 0.10000731
## x2 0.4611658883 -0.0595664742 0.12030340
## x3 0.5106209080 -0.0379672838 0.23151843
## x4 -0.3209858077 -0.1369755440 0.11303493
## x5 0.0486822084 -0.1592935097 0.13643980
## x7 -0.4112010606 -0.3421160933 -0.20291040
## x10 0.0009661917 -0.2136970776 -0.45713458
## x11 -0.0274064193 -0.0188026655 0.10903473
## x12 0.2651913926 0.2019457991 0.29815346
## x13 -0.3261039794 -0.2442275304 -0.13144087
## x14 -0.3145460320 -0.1799262766 -0.31860696
## x15 -0.0742246141 -0.0312169741 -0.02570451
## x16 -0.2632080095 0.0098798393 0.23149910
## x17 -0.2747539060 -0.0556453463 0.29196618
## x18 -0.4626325177 -0.2116118926 -0.11257880

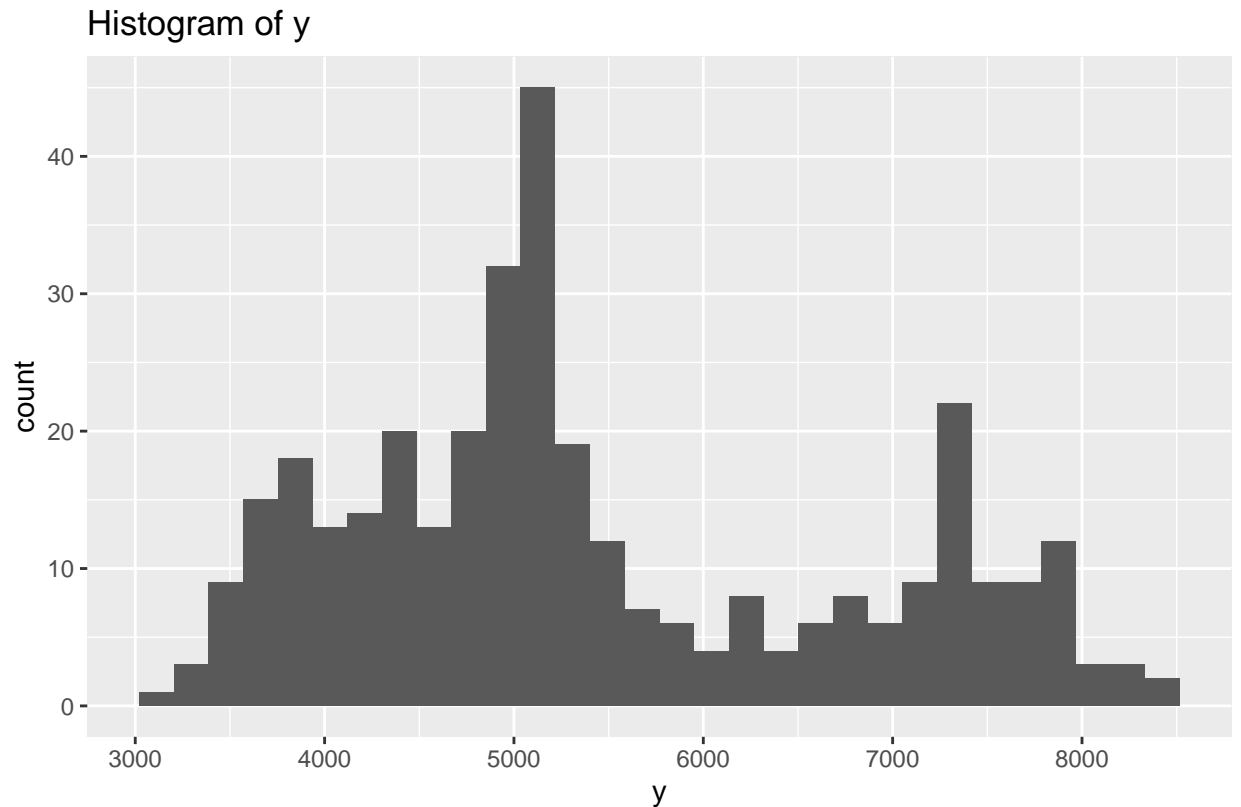
```

```
## x19 -0.0423140466 -0.0942398628 0.29756617
## x20 -0.1689090510 -0.0335161921 0.02455556
## x21 -0.1929635941 -0.0308827062 -0.03269500
## x23 -0.3261944813 0.0004736721 0.41090341
## x24 0.9117802988 0.2547417522 0.34580884
## x25 0.9070010098 0.4168428473 0.24615639
## x26 0.9339966376 0.3684401230 0.34027428
## x27 1.0000000000 0.3472926567 0.25472922
## x28 0.3472926567 1.0000000000 0.40875362
## x30 0.2547292197 0.4087536155 1.00000000
```

```
library(ggplot2)

# Create a histogram of the response variable y
ggplot(data, aes(x = y)) +
  geom_histogram() +
  labs(title = "Histogram of y",
       caption = "Data source: My Dataset")
```

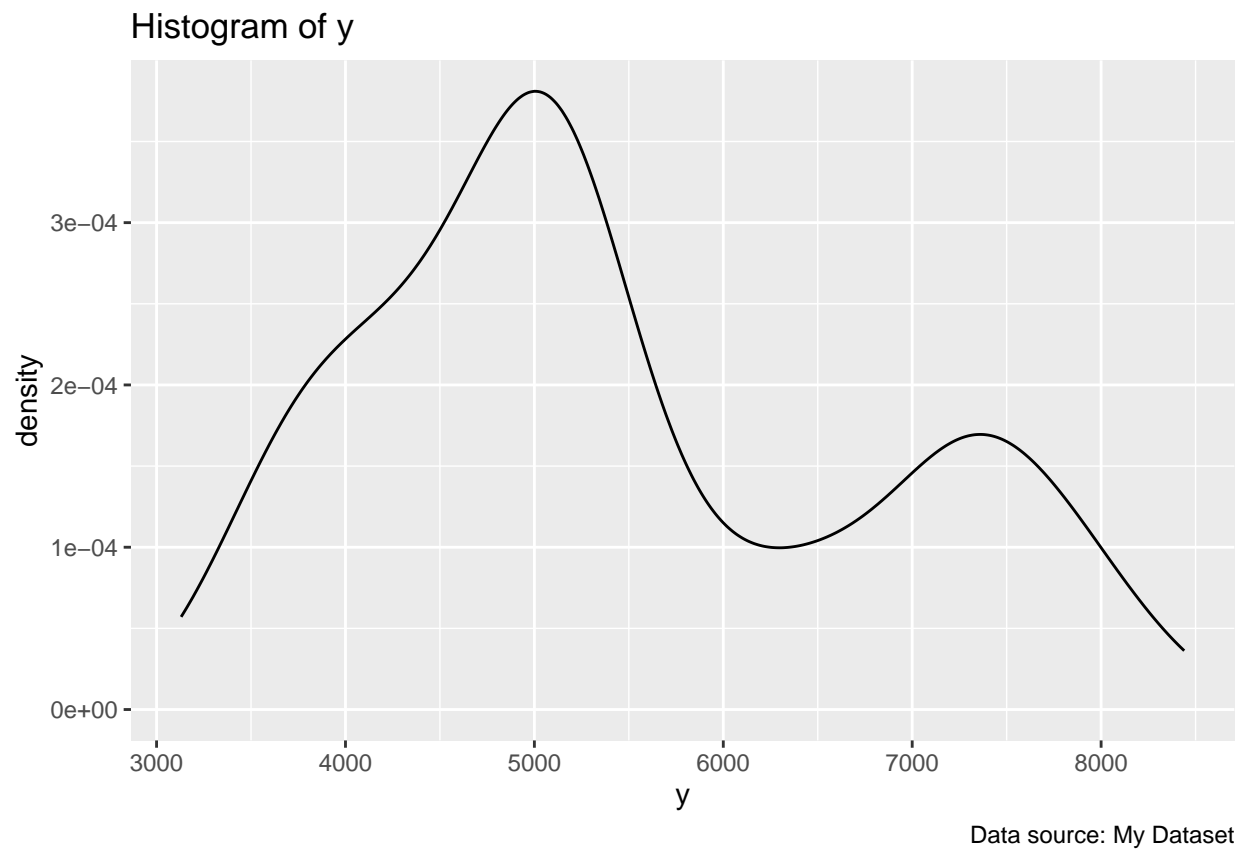
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Data source: My Dataset

```
# Create a density plot of the response variable y
ggplot(data, aes(x = y)) +
  geom_density() +
```

```
labs(title = "Histogram of y",
      caption = "Data source: My Dataset")
```

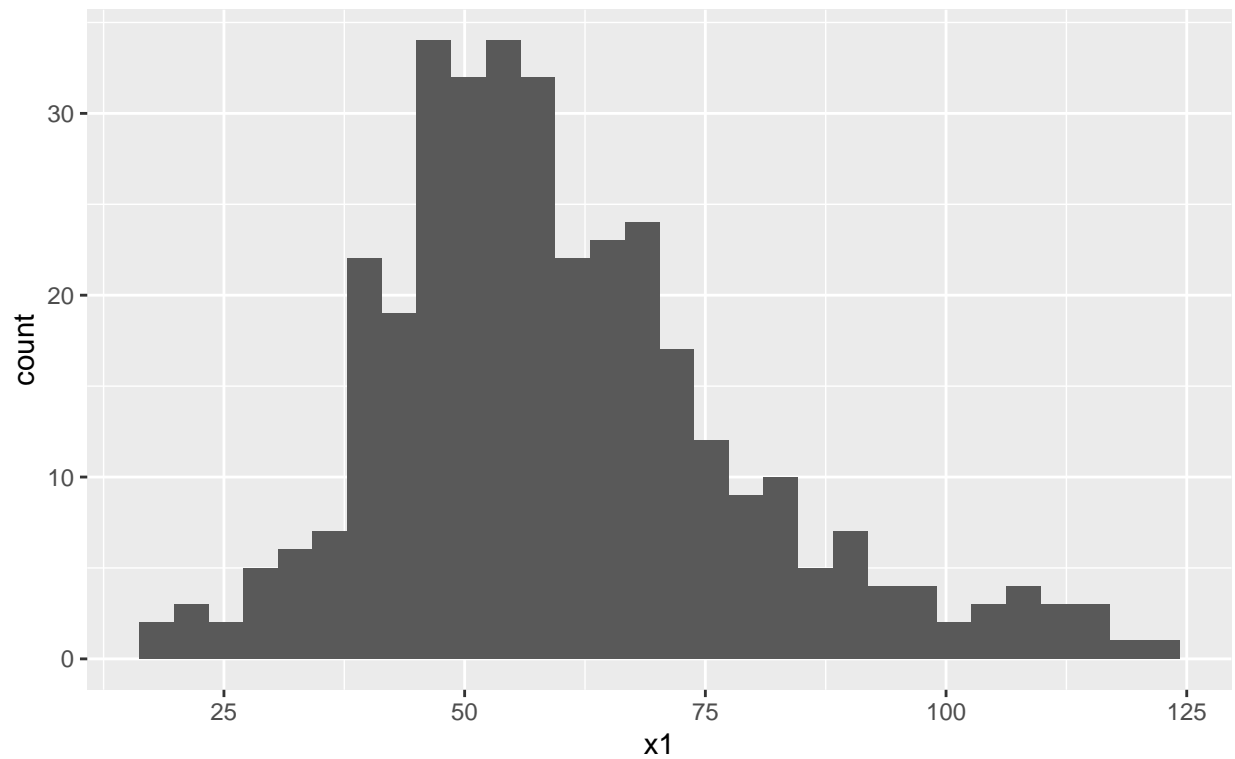


```
par(mfrow = c(3, 3))

# Create a histogram of the response variable x1
ggplot(data, aes(x = x1)) +
  geom_histogram() +
  labs(title = "Histogram of x1",
        caption = "Data source: My Dataset")
```

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

Histogram of x1

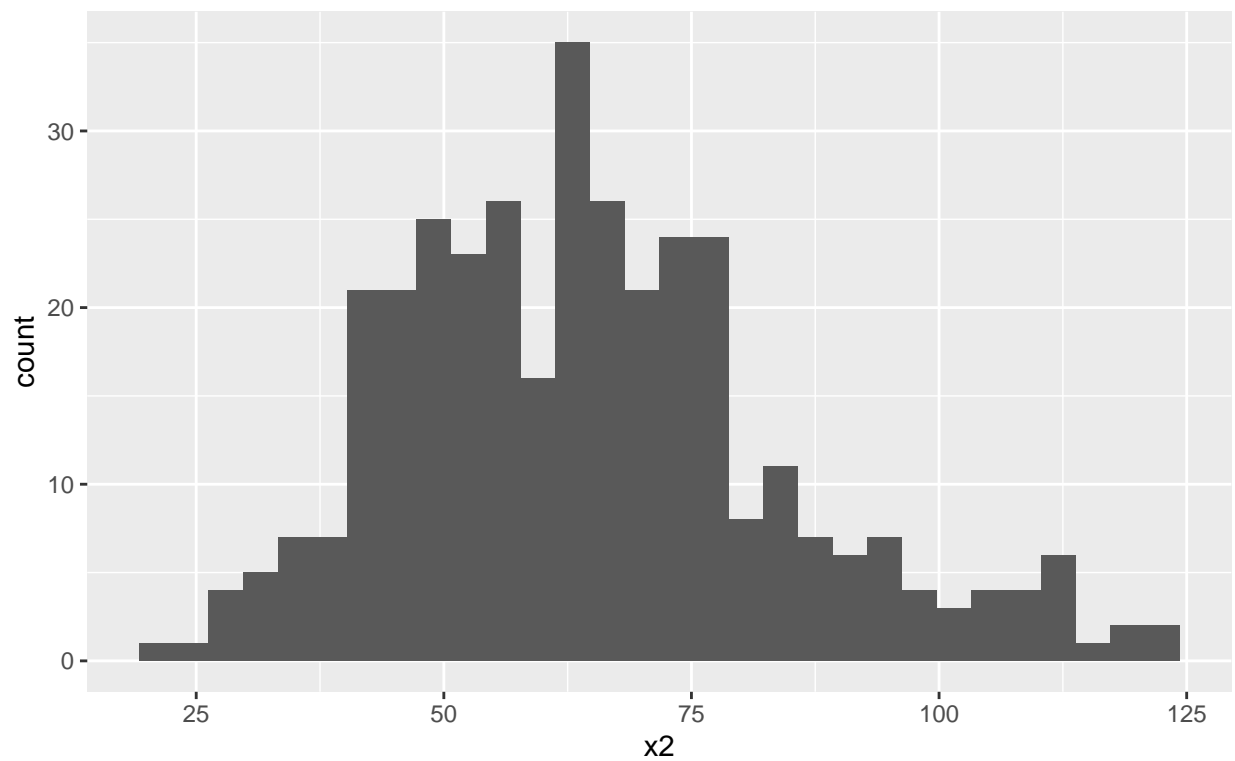


Data source: My Dataset

```
# Create a histogram of the response variable x2
ggplot(data, aes(x = x2)) +
  geom_histogram() +
  labs(title = "Histogram of x2",
        caption = "Data source: My Dataset")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```


Histogram of x2

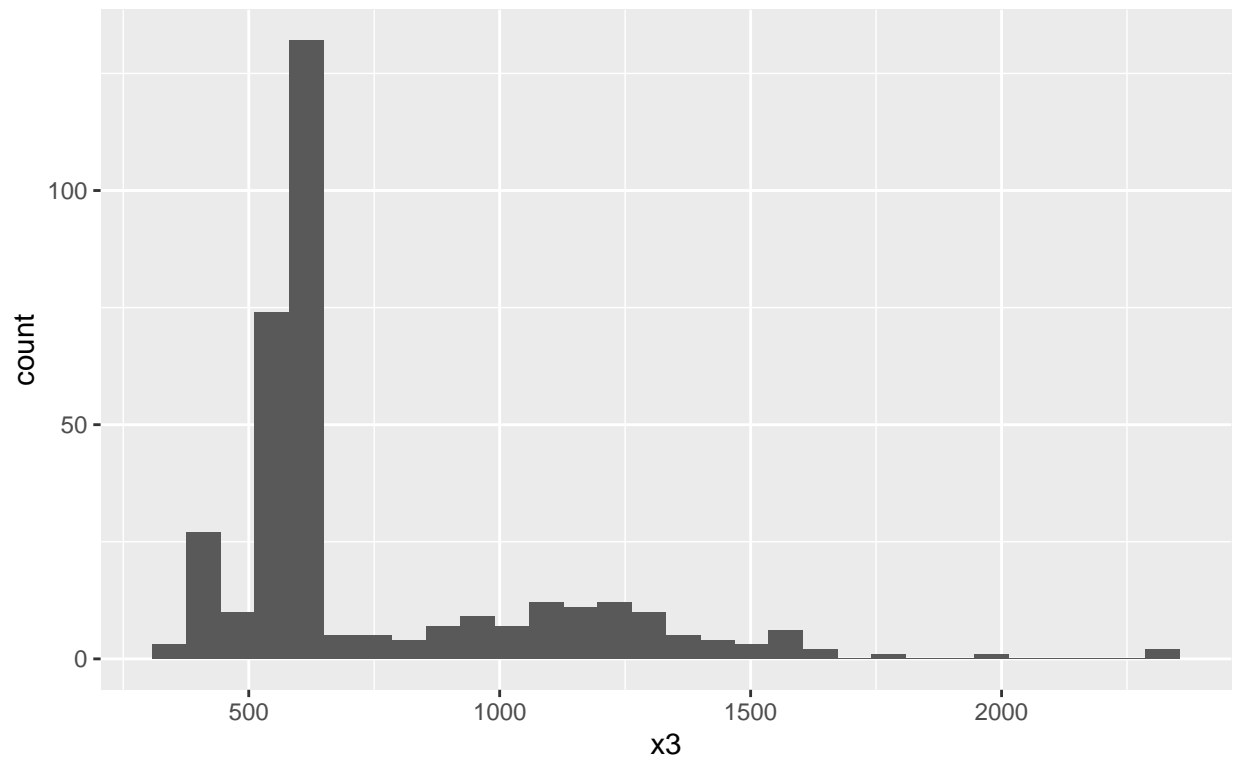


Data source: My Dataset

```
# Create a histogram of the response variable x3
ggplot(data, aes(x = x3)) +
  geom_histogram() +
  labs(title = "Histogram of x3",
        caption = "Data source: My Dataset")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

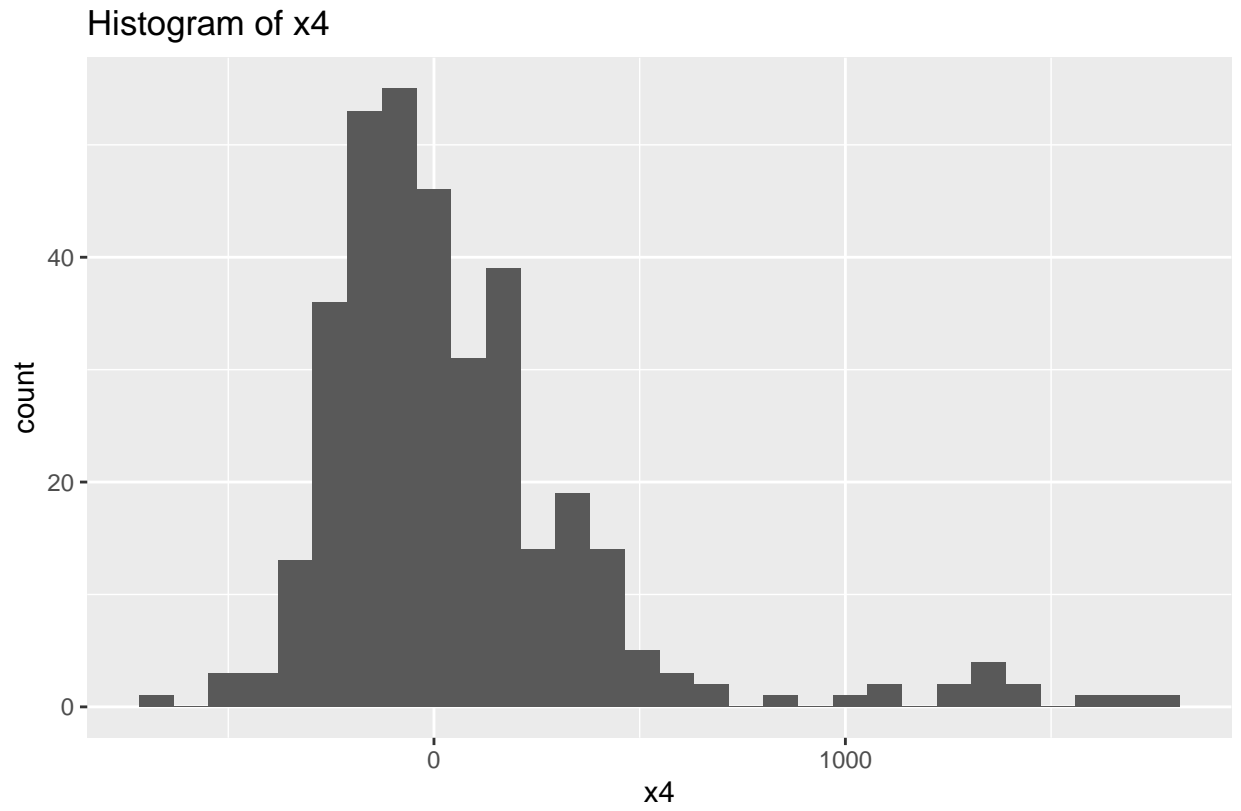
Histogram of x3



Data source: My Dataset

```
# Create a histogram of the response variable x4
ggplot(data, aes(x = x4)) +
  geom_histogram() +
  labs(title = "Histogram of x4",
       caption = "Data source: My Dataset")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



Data source: My Dataset