STAT 5243 Project 1

Code ▾

Hide

```
# Libraries
library(dplyr)
library(tidyr)
library(tidyverse)

# Loading dataset into R
setwd("~/Downloads")
```

```
Warning: The working directory was changed to /Users/safiaaladlouni/Downloads inside a n
otebook chunk. The working directory will be reset when the chunk is finished running. U
se the knitr root.dir option in the setup chunk to change the working directory for note
book chunks.
```

Hide

```
library(readxl)
Page1 <- read_excel("Page 1.xlsx", skip = 5) #skipping header rows
```

```
New names:
```

Hide

```
Page5 <- read_excel("Page 5.xlsx", skip = 3) #skipping header rows
```

Hide

```
#Running initial analyses to determine changes that need to be made to data

# Summary statistics
str(Page1)
```

```
tibble [7,435 × 97] (S3: tbl_df/tbl/data.frame)
 $ Plant Id                          : num [1:7435] 3 3 3 3 3 10 10 10 26 26 ...
 $ Combined Heat And
Power Plant      : chr [1:7435] "N" "N" "N" "N" ...
 $ Nuclear Unit Id                   : chr [1:7435] "." "." "." "." ...
 $ Plant Name                        : chr [1:7435] "Barry" "Barry" "Barry" "Barry" ...
 $ Operator Name                     : chr [1:7435] "Alabama Power Co" "Alabama Power C
o" "Alabama Power Co" "Alabama Power Co" ...
 $ Operator Id                       : num [1:7435] 195 195 195 195 195 195 195 195 195
195 ...
 $ Plant State                       : chr [1:7435] "AL" "AL" "AL" "AL" ...
 $ Census Region                     : chr [1:7435] "ESC" "ESC" "ESC" "ESC" ...
 $ NERC Region                       : chr [1:7435] "SERC" "SERC" "SERC" "SERC" ...
 $ Reserved...10                     : logi [1:7435] NA NA NA NA NA NA ...
 $ NAICS Code                        : num [1:7435] 22 22 22 22 22 22 22 22 22 22 ...
 $ EIA Sector Number                 : num [1:7435] 1 1 1 1 1 1 1 1 1 1 ...
 $ Sector Name                       : chr [1:7435] "Electric Utility" "Electric Utilit
y" "Electric Utility" "Electric Utility" ...
 $ Reported
Prime Mover        : chr [1:7435] "CA" "CT" "ST" "ST" ...
 $ Reported
Fuel Type Code        : chr [1:7435] "NG" "NG" "BIT" "NG" ...
 $ MER
Fuel Type Code            : chr [1:7435] "NG" "NG" "COL" "NG" ...
 $ Balancing
Authority Code     : chr [1:7435] "SOCO" "SOCO" "SOCO" "SOCO" ...
 $ Reserved...18                     : logi [1:7435] NA NA NA NA NA NA ...
 $ Physical
Unit Label        : chr [1:7435] "mcf" "mcf" "short tons" "mcf" ...
 $ Quantity
January           : chr [1:7435] "41646.0" "7803343.0" "44058.0" "679765.0" ...
 $ Quantity
February          : chr [1:7435] "19428.0" "7444935.0" "0.0" "0.0" ...
 $ Quantity
March             : chr [1:7435] "150437.0" "6575111.0" "0.0" "6.0" ...
 $ Quantity
April             : chr [1:7435] "184277.0" "5733373.0" "47328.0" "48911.0" ...
 $ Quantity
May               : chr [1:7435] "347065.0" "7124300.0" "0.0" "154483.0" ...
 $ Quantity
June              : chr [1:7435] "398051.0" "7384517.0" "44248.0" "335366.0" ...
 $ Quantity
July              : chr [1:7435] "538380.0" "7669669.0" "68612.0" "429006.0" ...
 $ Quantity
August            : chr [1:7435] "574143.0" "7658213.0" "0.0" "601460.0" ...
 $ Quantity
September         : chr [1:7435] "334012.0" "7423947.0" "15734.0" "47210.0" ...
 $ Quantity
October           : chr [1:7435] "231198.0" "7169534.0" "166757.0" "67868.0" ...
 $ Quantity
November          : chr [1:7435] "173270.0" "6679264.0" "8885.0" "11216.0" ...
 $ Quantity
```

```
December              : chr [1:7435] "." "." "." "." ...
 $ Elec_Quantity
January               : chr [1:7435] "41646.0" "7803343.0" "44058.0" "679765.0" ...
 $ Elec_Quantity
February              : chr [1:7435] "19428.0" "7444935.0" "0.0" "0.0" ...
 $ Elec_Quantity
March                 : chr [1:7435] "150437.0" "6575111.0" "0.0" "6.0" ...
 $ Elec_Quantity
April                 : chr [1:7435] "184277.0" "5733373.0" "47328.0" "48911.0" ...
 $ Elec_Quantity
May                   : chr [1:7435] "347065.0" "7124300.0" "0.0" "154483.0" ...
 $ Elec_Quantity
June                  : chr [1:7435] "398051.0" "7384517.0" "44248.0" "335366.0" ...
 $ Elec_Quantity
July                  : chr [1:7435] "538380.0" "7669669.0" "68612.0" "429006.0" ...
 $ Elec_Quantity
August                : chr [1:7435] "574143.0" "7658213.0" "0.0" "601460.0" ...
 $ Elec_Quantity
September             : chr [1:7435] "334012.0" "7423947.0" "15734.0" "47210.0" ...
 $ Elec_Quantity
October               : chr [1:7435] "231198.0" "7169534.0" "166757.0" "67868.0" ...
 $ Elec_Quantity
November              : chr [1:7435] "173270.0" "6679264.0" "8885.0" "11216.0" ...
 $ Elec_Quantity
December              : chr [1:7435] "." "." "." "." ...
 $ MMBtuPer_Unit
January               : chr [1:7435] "1.023" "1.023" "19.908" "1.023" ...
 $ MMBtuPer_Unit
February              : chr [1:7435] "1.023" "1.023" "0.0" "0.0" ...
 $ MMBtuPer_Unit
March                 : chr [1:7435] "1.021" "1.021" "0.0" "1.022" ...
 $ MMBtuPer_Unit
April                 : chr [1:7435] "1.018" "1.018" "19.908" "1.017" ...
 $ MMBtuPer_Unit
May                   : chr [1:7435] "1.015" "1.015" "0.0" "1.015" ...
 $ MMBtuPer_Unit
June                  : chr [1:7435] "1.019" "1.019" "19.908" "1.02" ...
 $ MMBtuPer_Unit
July                  : chr [1:7435] "1.022" "1.022" "19.908" "1.021" ...
 $ MMBtuPer_Unit
August                : chr [1:7435] "1.024" "1.024" "0.0" "1.023" ...
 $ MMBtuPer_Unit
September             : chr [1:7435] "1.023" "1.023" "19.243" "1.022" ...
 $ MMBtuPer_Unit
October               : chr [1:7435] "1.022" "1.022" "19.603" "1.023" ...
 $ MMBtuPer_Unit
November              : chr [1:7435] "1.018" "1.018" "19.485" "1.017" ...
 $ MMBtuPer_Unit
December              : chr [1:7435] "." "." "." "." ...
 $ Tot_MMBtu
January               : chr [1:7435] "42604.0" "7982820.0" "877107.0" "695400.0" ...
 $ Tot_MMBtu
```

```
 February                : chr [1:7435] "19875.0" "7616169.0" "0.0" "0.0" ...
 $ Tot_MMBtu
 March                   : chr [1:7435] "153596.0" "6713188.0" "0.0" "6.0" ...
 $ Tot_MMBtu
 April                   : chr [1:7435] "187594.0" "5836574.0" "942206.0" "49742.0" ...
 $ Tot_MMBtu
 May                     : chr [1:7435] "352271.0" "7231165.0" "0.0" "156800.0" ...
 $ Tot_MMBtu
 June                    : chr [1:7435] "405614.0" "7524823.0" "880889.0" "342073.0" ...
 $ Tot_MMBtu
 July                    : chr [1:7435] "550224.0" "7838402.0" "1365928.0" "438015.0" ...
 $ Tot_MMBtu
 August                  : chr [1:7435] "587922.0" "7842010.0" "0.0" "615294.0" ...
 $ Tot_MMBtu
 September               : chr [1:7435] "341694.0" "7594698.0" "302769.0" "48249.0" ...
 $ Tot_MMBtu
 October                 : chr [1:7435] "236284.0" "7327264.0" "3268937.0" "69429.0" ...
 $ Tot_MMBtu
 November                : chr [1:7435] "176389.0" "6799491.0" "173124.0" "11407.0" ...
 $ Tot_MMBtu
 December                : chr [1:7435] "." "." "." "." ...
 $ Elec_MMBtu
 January                 : chr [1:7435] "42604.0" "7982820.0" "877107.0" "695400.0" ...
 $ Elec_MMBtu
 February                : chr [1:7435] "19875.0" "7616169.0" "0.0" "0.0" ...
 $ Elec_MMBtu
 March                   : chr [1:7435] "153596.0" "6713188.0" "0.0" "6.0" ...
 $ Elec_MMBtu
 April                   : chr [1:7435] "187594.0" "5836574.0" "942206.0" "49742.0" ...
 $ Elec_MMBtu
 May                     : chr [1:7435] "352271.0" "7231165.0" "0.0" "156800.0" ...
 $ Elec_MMBtu
 June                    : chr [1:7435] "405614.0" "7524823.0" "880889.0" "342073.0" ...
 $ Elec_MMBtu
 July                    : chr [1:7435] "550224.0" "7838402.0" "1365928.0" "438015.0" ...
 $ Elec_MMBtu
 August                  : chr [1:7435] "587922.0" "7842010.0" "0.0" "615294.0" ...
 $ Elec_MMBtu
 September               : chr [1:7435] "341694.0" "7594698.0" "302769.0" "48249.0" ...
 $ Elec_MMBtu
 October                 : chr [1:7435] "236284.0" "7327264.0" "3268937.0" "69429.0" ...
 $ Elec_MMBtu
 November                : chr [1:7435] "176389.0" "6799491.0" "173124.0" "11407.0" ...
 $ Elec_MMBtu
 December                : chr [1:7435] "." "." "." "." ...
 $ Netgen
 January                     : chr [1:7435] "397842.0" "791439.0" "74699.628" "59224.372"
 ...
 $ Netgen
 February                    : chr [1:7435] "381637.0" "754359.0" "-4538.357" "-1664.643"
 ...
 $ Netgen
```

```
 March                       : chr [1:7435] "342909.0" "685854.0" "0.0" "-4938.0" ...
  $ Netgen
 April                       : chr [1:7435] "301512.0" "583917.0" "83997.467" "4434.533"
 ...
  $ Netgen
 May                         : chr [1:7435] "389596.0" "723389.0" "0.0" "5820.0" ...
  $ Netgen
 June                        : chr [1:7435] "405885.0" "755776.0" "75098.285" "29162.715"
 ...
  $ Netgen
 July                        : chr [1:7435] "433857.0" "783231.0" "114680.23" "36774.769"
 ...
  $ Netgen
 August                      : chr [1:7435] "433183.0" "789596.0" "0.0" "34781.0" ...
  $ Netgen
 September                   : chr [1:7435] "400744.0" "762657.0" "20325.048" "3238.952"
 ...
  $ Netgen
 October                     : chr [1:7435] "380868.0" "735257.0" "322168.46" "6842.536"
 ...
  $ Netgen
 November                    : chr [1:7435] "346913.0" "675220.0" "11097.797" "731.203"
 ...
  $ Netgen
 December                    : chr [1:7435] "." "." "." "." ...
  $ Total Fuel Consumption
 Quantity   : num [1:7435] 2991907 78666206 395622 2375291 0 ...
  $ Electric Fuel Consumption
 Quantity: num [1:7435] 2991907 78666206 395622 2375291 0 ...
  $ Total Fuel Consumption
 MMBtu       : num [1:7435] 3054067 80306604 7810960 2426415 0 ...
  $ Elec Fuel Consumption
 MMBtu        : num [1:7435] 3054067 80306604 7810960 2426415 0 ...
  $ Net Generation
 (Megawatthours)     : num [1:7435] 4214946 8040695 697529 174407 0 ...
  $ YEAR                        : num [1:7435] 2024 2024 2024 2024 2024 ...
```

Hide

```
summary(Page1)
```

```
   Plant Id      Combined Heat And\nPower Plant Nuclear Unit Id     Plant Name         Op
erator Name       Operator Id
 Min.   :    3   Length:7435                     Length:7435      Length:7435        Le
ngth:7435       Min.   :    8
 1st Qu.:10430   Class :character                Class :character Class :character   Cl
ass :character   1st Qu.:13598
 Median :57210   Mode  :character                Mode  :character Mode  :character   Mo
de  :character   Median :56215
 Mean   :53038
Mean   :47510
 3rd Qu.:66188
3rd Qu.:65411
 Max.   :99999
Max.   :99999
 Plant State        Census Region       NERC Region       Reserved...10    NAICS Code
EIA Sector Number Sector Name
 Length:7435        Length:7435        Length:7435       Mode:logical   Min.   :    22
Min.   :1.000     Length:7435
 Class :character   Class :character   Class :character  NA's:7435      1st Qu.:    22
1st Qu.:1.000     Class :character
 Mode  :character   Mode  :character   Mode  :character                 Median :    22
Median :2.000     Mode  :character
                                                                        Mean   : 38548
Mean   :2.664
                                                                        3rd Qu.: 99999
3rd Qu.:3.000
                                                                        Max.   :562213
Max.   :7.000
 Reported\nPrime Mover Reported\nFuel Type Code MER\nFuel Type Code Balancing\nAuthority
Code Reserved...18
 Length:7435           Length:7435             Length:7435          Length:7435
Mode:logical
 Class :character      Class :character        Class :character     Class :character
NA's:7435
 Mode  :character      Mode  :character        Mode  :character     Mode  :character




 Physical\nUnit Label Quantity\nJanuary  Quantity\nFebruary Quantity\nMarch    Quantity
\nApril    Quantity\nMay
 Length:7435          Length:7435         Length:7435        Length:7435        Length:74
35        Length:7435
 Class :character     Class :character    Class :character   Class :character   Class :ch
aracter   Class :character
 Mode  :character     Mode  :character    Mode  :character   Mode  :character   Mode  :ch
aracter   Mode  :character




 Quantity\nJune      Quantity\nJuly      Quantity\nAugust    Quantity\nSeptember Quantity\n
October  Quantity\nNovember
 Length:7435         Length:7435         Length:7435         Length:7435         Length:743
```

```
 5          Length:7435
 Class :character   Class :character   Class :character   Class :character    Class :cha
racter    Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character    Mode  :cha
racter    Mode  :character




 Quantity\nDecember Elec_Quantity\nJanuary Elec_Quantity\nFebruary Elec_Quantity\nMarch
Elec_Quantity\nApril Elec_Quantity\nMay
 Length:7435         Length:7435          Length:7435          Length:7435
Length:7435          Length:7435
 Class :character    Class :character     Class :character     Class :character
Class :character     Class :character
 Mode  :character    Mode  :character     Mode  :character     Mode  :character
Mode  :character     Mode  :character




 Elec_Quantity\nJune Elec_Quantity\nJuly Elec_Quantity\nAugust Elec_Quantity\nSeptember
Elec_Quantity\nOctober
 Length:7435         Length:7435          Length:7435          Length:7435
Length:7435
 Class :character    Class :character     Class :character     Class :character
Class :character
 Mode  :character    Mode  :character     Mode  :character     Mode  :character
Mode  :character




 Elec_Quantity\nNovember Elec_Quantity\nDecember MMBtuPer_Unit\nJanuary MMBtuPer_Unit\nF
ebruary MMBtuPer_Unit\nMarch
 Length:7435             Length:7435             Length:7435             Length:7435
Length:7435
 Class :character        Class :character        Class :character        Class :character
Class :character
 Mode  :character        Mode  :character        Mode  :character        Mode  :character
Mode  :character




 MMBtuPer_Unit\nApril MMBtuPer_Unit\nMay MMBtuPer_Unit\nJune MMBtuPer_Unit\nJuly MMBtuPe
r_Unit\nAugust MMBtuPer_Unit\nSeptember
 Length:7435          Length:7435         Length:7435         Length:7435         Length:
7435            Length:7435
 Class :character     Class :character    Class :character    Class :character    Class :
character        Class :character
 Mode  :character     Mode  :character    Mode  :character    Mode  :character    Mode  :
character        Mode  :character
```

```
 MMBtuPer_Unit\nOctober MMBtuPer_Unit\nNovember MMBtuPer_Unit\nDecember Tot_MMBtu\nJanua
ry Tot_MMBtu\nFebruary
 Length:7435            Length:7435             Length:7435             Length:7435
Length:7435
 Class :character        Class :character        Class :character        Class :character
Class :character
 Mode  :character        Mode  :character        Mode  :character        Mode  :character
Mode  :character
```

```
 Tot_MMBtu\nMarch    Tot_MMBtu\nApril    Tot_MMBtu\nMay     Tot_MMBtu\nJune    Tot_MMBtu\n
July    Tot_MMBtu\nAugust
 Length:7435        Length:7435        Length:7435        Length:7435        Length:7435
Length:7435
 Class :character    Class :character    Class :character    Class :character    Class :char
acter    Class :character
 Mode  :character    Mode  :character    Mode  :character    Mode  :character    Mode  :char
acter    Mode  :character
```

```
 Tot_MMBtu\nSeptember Tot_MMBtu\nOctober Tot_MMBtu\nNovember Tot_MMBtu\nDecember Elec_MM
Btu\nJanuary Elec_MMBtu\nFebruary
 Length:7435            Length:7435         Length:7435         Length:7435         Length:
7435         Length:7435
 Class :character        Class :character    Class :character    Class :character    Class :
character    Class :character
 Mode  :character        Mode  :character    Mode  :character    Mode  :character    Mode  :
character    Mode  :character
```

```
 Elec_MMBtu\nMarch  Elec_MMBtu\nApril  Elec_MMBtu\nMay     Elec_MMBtu\nJune    Elec_MMBtu
\nJuly    Elec_MMBtu\nAugust
 Length:7435        Length:7435        Length:7435        Length:7435        Length:7435
Length:7435
 Class :character    Class :character    Class :character    Class :character    Class :char
acter    Class :character
 Mode  :character    Mode  :character    Mode  :character    Mode  :character    Mode  :char
acter    Mode  :character
```

```
 Elec_MMBtu\nSeptember Elec_MMBtu\nOctober Elec_MMBtu\nNovember Elec_MMBtu\nDecember Net
gen\nJanuary    Netgen\nFebruary
 Length:7435        Length:7435         Length:7435          Length:7435          Len
gth:7435         Length:7435
 Class :character        Class :character    Class :character    Class :character        Cla
ss :character    Class :character
 Mode  :character        Mode  :character    Mode  :character    Mode  :character        Mod
e :character    Mode  :character
```

```
 Netgen\nMarch        Netgen\nApril         Netgen\nMay          Netgen\nJune         Netgen\nJu
ly       Netgen\nAugust
 Length:7435          Length:7435          Length:7435          Length:7435          Length:7435
Length:7435
 Class :character     Class :character     Class :character     Class :character     Class :char
acter    Class :character
 Mode  :character     Mode  :character     Mode  :character     Mode  :character     Mode  :char
acter    Mode  :character



 Netgen\nSeptember   Netgen\nOctober      Netgen\nNovember     Netgen\nDecember     Total Fuel
Consumption\nQuantity
 Length:7435          Length:7435          Length:7435          Length:7435          Min.   :
0
 Class :character     Class :character     Class :character     Class :character     1st Qu.:
0
 Mode  :character     Mode  :character     Mode  :character     Mode  :character     Median :
0
                                                                                      Mean   :  2
068802
                                                                                      3rd Qu.:
278867
                                                                                      Max.   :167
615627
 Electric Fuel Consumption\nQuantity Total Fuel Consumption\nMMBtu Elec Fuel Consumption
\nMMBtu Net Generation\n(Megawatthours)
 Min.   :       0                    Min.   :       0                  Min.   :       0
Min.   :-1025528
 1st Qu.:       0                    1st Qu.:       0                  1st Qu.:       0
1st Qu.:       8
 Median :       0                    Median :   234311                 Median :   159464
Median :   34890
 Mean   : 1823200                    Mean   :  4425718                 Mean   :  4119720
Mean   :  529998
 3rd Qu.:  156770                    3rd Qu.:  1797345                 3rd Qu.:  1395478
3rd Qu.:  353670
 Max.   :130115041                   Max.   :170637388                 Max.   :170637388
Max.   :26265163
      YEAR
 Min.   :2024
 1st Qu.:2024
 Median :2024
 Mean   :2024
 3rd Qu.:2024
 Max.   :2024
```

Hide

```
str(Page5)
```

```
tibble [20,779 × 31] (S3: tbl_df/tbl/data.frame)
 $ YEAR                             : num [1:20779] 2024 2024 2024 2024 2024 ...
 $ MONTH                            : num [1:20779] 1 1 1 1 1 1 1 1 1 1 ...
 $ Plant Id                         : num [1:20779] 3 3 3 10 26 26 26 54 56 56 ...
 $ Plant Name                       : chr [1:20779] "Barry" "Barry" "Barry" "Greene Cou
nty" ...
 $ Plant State                      : chr [1:20779] "AL" "AL" "AL" "AL" ...
 $ Purchase Type                    : chr [1:20779] "C" "S" "S" "S" ...
 $ Contract
Expiration Date        : chr [1:20779] "624.0" "." "." "." ...
 $ ENERGY_SOURCE                    : chr [1:20779] "BIT" "NG" "NG" "NG" ...
 $ FUEL_GROUP                       : chr [1:20779] "Coal" "Natural Gas" "Natural Gas"
"Natural Gas" ...
 $ Coalmine
Type                   : chr [1:20779] "S" NA NA NA ...
 $ Coalmine
State                  : chr [1:20779] "CL" NA NA NA ...
 $ Coalmine
County                 : chr [1:20779] "IMP" NA NA NA ...
 $ Coalmine
Msha Id                : chr [1:20779] "NA" NA NA NA ...
 $ Coalmine
Name                   : chr [1:20779] "MINA PRIBBENOW" NA NA NA ...
 $ SUPPLIER                         : chr [1:20779] "INTEROCEAN COAL SALES" "VARIOUS (N
ATURAL GAS SPOT PURCHASES ONLY)" "VARIOUS (NATURAL GAS SPOT PURCHASES ONLY)" "VARIOUS (N
ATURAL GAS SPOT PURCHASES ONLY)" ...
 $ QUANTITY                         : num [1:20779] 65880 679779 6759960 544609 63686
...
 $ Average Heat
Content           : num [1:20779] 20.31 1.02 1.02 1.03 24.44 ...
 $ Average Sulfur
Content         : num [1:20779] 0.75 0 0 0 1.21 2.21 0 0 0 0 ...
 $ Average Ash
Content            : num [1:20779] 11.8 0 0 0 13.6 11.1 0 0 0 0 ...
 $ Average Mercury
Content        : num [1:20779] 0 0 0 0 0 0 0 0 0 0 ...
 $ FUEL_COST                        : chr [1:20779] "548.0" "373.8" "376.0" "671.2" ...
 $ Regulated                        : chr [1:20779] "REG" "REG" "REG" "REG" ...
 $ Operator Name                    : chr [1:20779] "Alabama Power Co" "Alabama Power C
o" "Alabama Power Co" "Alabama Power Co" ...
 $ Operator Id                      : num [1:20779] 195 195 195 195 195 195 195 5580 18
9 189 ...
 $ Primary Transportation Mode      : chr [1:20779] "RV" "PL" "PL" "PL" ...
 $ Secondary Transportation Mode    : chr [1:20779] NA NA NA NA ...
 $ Natural Gas Supply Contract Type : chr [1:20779] NA "F" "F" "F" ...
 $ Natural Gas Delivery Contract Type: chr [1:20779] NA "F" "F" "F" ...
 $ Moisture
Content                : chr [1:20779] "16.0" "." "." "." ...
 $ Chlorine
Content                : chr [1:20779] "0.0" "." "." "." ...
 $ BA_CODE                          : chr [1:20779] "SOCO" "SOCO" "SOCO" "SOCO" ...
```

```
summary(Page5)
```

```
      YEAR            MONTH          Plant Id        Plant Name         Plant State        P
urchase Type
 Min.   :2024    Min.   : 1.000   Min.   :    3   Length:20779      Length:20779         L
ength:20779
 1st Qu.:2024    1st Qu.: 3.000   1st Qu.: 2963    Class :character   Class :character    C
lass :character
 Median :2024    Median : 6.000   Median : 7757    Mode  :character   Mode  :character    M
ode  :character
 Mean   :2024    Mean   : 5.952   Mean   :22889
 3rd Qu.:2024    3rd Qu.: 9.000   3rd Qu.:55298
 Max.   :2024    Max.   :11.000   Max.   :67766
 Contract\nExpiration Date ENERGY_SOURCE      FUEL_GROUP        Coalmine\nType     Coal
mine\nState    Coalmine\nCounty
 Length:20779            Length:20779       Length:20779       Length:20779       Leng
th:20779       Length:20779
 Class :character         Class :character  Class :character   Class :character   Clas
s :character   Class :character
 Mode  :character         Mode  :character  Mode  :character   Mode  :character   Mode
:character    Mode  :character



 Coalmine\nMsha Id  Coalmine\nName      SUPPLIER          QUANTITY         Average Hea
t\nContent Average Sulfur\nContent
 Length:20779       Length:20779       Length:20779     Min.   :       1  Min.   : 0.
833      Min.   :0.0000
 Class :character   Class :character   Class :character  1st Qu.:    15618  1st Qu.: 1.
028       1st Qu.:0.0000
 Mode  :character   Mode  :character   Mode  :character  Median :    75021  Median : 1.
041       Median :0.0000
                                                         Mean   :   573183  Mean   : 6.
684       Mean   :0.3807
                                                         3rd Qu.:   508556  3rd Qu.:16.
780       3rd Qu.:0.2100
                                                         Max.   :15263613  Max.   :29.
134       Max.   :8.4400
 Average Ash\nContent Average Mercury\nContent  FUEL_COST        Regulated          Ope
rator Name       Operator Id
 Min.   : 0.000      Min.   :0.000000    Length:20779       Length:20779       Len
gth:20779      Min.   :   35
 1st Qu.: 0.000      1st Qu.:0.000000    Class :character   Class :character   Cla
ss :character  1st Qu.: 6455
 Median : 0.000      Median :0.000000    Mode  :character   Mode  :character   Mod
e :character   Median :13994
 Mean   : 2.188      Mean   :0.004377                                          Mean   :18752
 3rd Qu.: 4.400      3rd Qu.:0.000000                                          3rd Qu.:18642
 Max.   :63.400      Max.   :0.571000                                          Max.   :66103
 Primary Transportation Mode Secondary Transportation Mode Natural Gas Supply Contract T
ype Natural Gas Delivery Contract Type
```

```
  Length:20779                    Length:20779                    Length:20779
 Length:20779
  Class :character                Class :character                Class :character
 Class :character
  Mode  :character                Mode  :character                Mode  :character
 Mode  :character



 Moisture\nContent  Chlorine\nContent     BA_CODE
 Length:20779       Length:20779       Length:20779
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character
```

Hide

```
# Checking column names and determining how data is being organized
colnames(Page1)
```

```
 [1] "Plant Id"                         "Combined Heat And\nPower Plant"      "Nuclea
r Unit Id"
 [4] "Plant Name"                       "Operator Name"                      "Operat
or Id"
 [7] "Plant State"                      "Census Region"                      "NERC R
egion"
[10] "Reserved...10"                    "NAICS Code"                         "EIA Se
ctor Number"
[13] "Sector Name"                      "Reported\nPrime Mover"              "Report
ed\nFuel Type Code"
[16] "MER\nFuel Type Code"              "Balancing\nAuthority Code"          "Reserv
ed...18"
[19] "Physical\nUnit Label"            "Quantity\nJanuary"                   "Quanti
ty\nFebruary"
[22] "Quantity\nMarch"                  "Quantity\nApril"                    "Quanti
ty\nMay"
[25] "Quantity\nJune"                   "Quantity\nJuly"                     "Quanti
ty\nAugust"
[28] "Quantity\nSeptember"             "Quantity\nOctober"                   "Quanti
ty\nNovember"
[31] "Quantity\nDecember"              "Elec_Quantity\nJanuary"             "Elec_Q
uantity\nFebruary"
[34] "Elec_Quantity\nMarch"            "Elec_Quantity\nApril"               "Elec_Q
uantity\nMay"
[37] "Elec_Quantity\nJune"             "Elec_Quantity\nJuly"                "Elec_Q
uantity\nAugust"
[40] "Elec_Quantity\nSeptember"        "Elec_Quantity\nOctober"             "Elec_Q
uantity\nNovember"
[43] "Elec_Quantity\nDecember"         "MMBtuPer_Unit\nJanuary"             "MMBtuP
er_Unit\nFebruary"
[46] "MMBtuPer_Unit\nMarch"            "MMBtuPer_Unit\nApril"               "MMBtuP
er_Unit\nMay"
[49] "MMBtuPer_Unit\nJune"             "MMBtuPer_Unit\nJuly"                "MMBtuP
er_Unit\nAugust"
[52] "MMBtuPer_Unit\nSeptember"        "MMBtuPer_Unit\nOctober"             "MMBtuP
er_Unit\nNovember"
[55] "MMBtuPer_Unit\nDecember"         "Tot_MMBtu\nJanuary"                 "Tot_MM
Btu\nFebruary"
[58] "Tot_MMBtu\nMarch"                "Tot_MMBtu\nApril"                   "Tot_MM
Btu\nMay"
[61] "Tot_MMBtu\nJune"                 "Tot_MMBtu\nJuly"                    "Tot_MM
Btu\nAugust"
[64] "Tot_MMBtu\nSeptember"            "Tot_MMBtu\nOctober"                 "Tot_MM
Btu\nNovember"
[67] "Tot_MMBtu\nDecember"             "Elec_MMBtu\nJanuary"                "Elec_M
MBtu\nFebruary"
[70] "Elec_MMBtu\nMarch"               "Elec_MMBtu\nApril"                  "Elec_M
MBtu\nMay"
[73] "Elec_MMBtu\nJune"                "Elec_MMBtu\nJuly"                   "Elec_M
MBtu\nAugust"
[76] "Elec_MMBtu\nSeptember"           "Elec_MMBtu\nOctober"                "Elec_M
MBtu\nNovember"
```

```
 [79] "Elec_MMBtu\nDecember"          "Netgen\nJanuary"                    "Netgen
\nFebruary"
 [82] "Netgen\nMarch"                 "Netgen\nApril"                      "Netgen
\nMay"
 [85] "Netgen\nJune"                  "Netgen\nJuly"                       "Netgen
\nAugust"
 [88] "Netgen\nSeptember"            "Netgen\nOctober"                    "Netgen
\nNovember"
 [91] "Netgen\nDecember"             "Total Fuel Consumption\nQuantity"   "Electr
ic Fuel Consumption\nQuantity"
 [94] "Total Fuel Consumption\nMMBtu" "Elec Fuel Consumption\nMMBtu"       "Net Ge
neration\n(Megawatthours)"
 [97] "YEAR"
```

<div align="right">Hide</div>

```
colnames(Page5)
```

```
  [1] "YEAR"                                  "MONTH"                          "Plant I
d"
  [4] "Plant Name"                            "Plant State"                    "Purchase
Type"
  [7] "Contract\nExpiration Date"             "ENERGY_SOURCE"                  "FUEL_GRO
UP"
 [10] "Coalmine\nType"                        "Coalmine\nState"                "Coalmine
\nCounty"
 [13] "Coalmine\nMsha Id"                     "Coalmine\nName"                 "SUPPLIE
R"
 [16] "QUANTITY"                              "Average Heat\nContent"          "Average
Sulfur\nContent"
 [19] "Average Ash\nContent"                  "Average Mercury\nContent"       "FUEL_COS
T"
 [22] "Regulated"                             "Operator Name"                  "Operator
Id"
 [25] "Primary Transportation Mode"           "Secondary Transportation Mode"  "Natural
Gas Supply Contract Type"
 [28] "Natural Gas Delivery Contract Type" "Moisture\nContent"                 "Chlorine
\nContent"
 [31] "BA_CODE"
```

<div align="right">Hide</div>

```
# Checking for NAs
table(is.na(Page1))
```

```
 FALSE    TRUE
702170   19025
```

<div align="right">Hide</div>

```
table(is.na(Page5))
```

```
  FALSE    TRUE
 533194  110955
```

<div style="text-align: right">Hide</div>

```
# In page 1, we want to consider the following variables for Generation and Fuel Data:
[Month, Quantity Consumed in Physical Units for Electric Generation, Physical Unit Labe
l, Reported Prime Mover, Reported Fuel Type Code, Operator Name (or ID), Plant State]
# In page 5, we want to consider the following variables: [ENERGY_SOURCE, FUEL_GROUP, QU
ANTITY, Average Heat Content, Average Sulfur Content, Average Ash Content, FUEL_COST, Mo
isture Content, Primary Transportation Mode]
```

<div style="text-align: right">Hide</div>

```
# Renaming columns for consistency and pivoting
Page1 <- Page1 %>%
  rename(
    `Operator ID` = `Operator Id`,
    `Plant ID` = `Plant Id`,
    `Nuclear Unit ID` = `Nuclear Unit Id`,
    `Year` = `YEAR`,
    `Reported Prime Mover` = `Reported\nPrime Mover`,
    `ReportedFuelTypeCode` = `Reported\nFuel Type Code`,
    `Physical Unit Label` = `Physical\nUnit Label`
  ) %>%
    pivot_longer(
    cols = starts_with("Quantity\n"),
    names_to = "Month",
    values_to = "Quantity Consumed in Physical Units for Electric Generation"
    ) %>%
    mutate(
    Month = gsub("Quantity\n", "", Month) # Renaming month names
    )

# Modified dataframe:
head(Page1)
```

Plant ID	Combined Heat And Power Plant	Nuclear Unit ID	Plant Name	Operator Name
<dbl>	<chr>	<chr>	<chr>	<chr>
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co

Plant ID <dbl>	Combined Heat And Power Plant <chr>	Nuclear Unit ID <chr>	Plant Name <chr>	Operator Name <chr>
3	N	.	Barry	Alabama Power Co

6 rows | 1-6 of 87 columns

<div style="text-align: right;">Hide</div>

```
Page1 <- Page1 %>%
  mutate(Month = match(Month, month.name))

head(Page1)
```

Plant ID <dbl>	Combined Heat And Power Plant <chr>	Nuclear Unit ID <chr>	Plant Name <chr>	Operator Name <chr>
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co
3	N	.	Barry	Alabama Power Co

6 rows | 1-6 of 87 columns

<div style="text-align: right;">Hide</div>

```
# Renaming columns for consistency and selecting variables of interest

unique(Page5$Month)
```

```
Warning: Unknown or uninitialised column: `Month`.
```

```
NULL
```

<div style="text-align: right;">Hide</div>

```r
Page5 <- Page5 %>%
  rename(
    `Average Heat Content` = `Average Heat\nContent`,
    `Average Sulfur Content` = `Average Sulfur\nContent`,
    `Average Ash Content` = `Average Ash\nContent`,
    `Contract Expiration Date` = `Contract\nExpiration Date`,
    `Coalmine Type` = `Coalmine\nType`,
    `Coalmine State` = `Coalmine\nState`,
    `Coalmine County` = `Coalmine\nCounty`,
    `Coalmine Name` = `Coalmine\nName`,
    `Coalmine Msha ID` = `Coalmine\nMsha Id`,
    `Average Mercury Content` = `Average Mercury\nContent`,
    `Moisture Content` = `Moisture\nContent`,
    `Chlorine Content` = `Chlorine\nContent`,
    `EnergySource` = `ENERGY_SOURCE`,
    `Fuel Group` = `FUEL_GROUP`,
    `Quantity` = `QUANTITY`,
    `Year` = `YEAR`,
    `Month` = `MONTH`,
    `Plant ID` = `Plant Id`,
    `Supplier` = `SUPPLIER`,
    `Operator ID` = `Operator Id`,
    `Fuel Cost` = `FUEL_COST`
  ) %>%
  select(
    `EnergySource`,
    `Fuel Group`,
    `Quantity`,
    `Average Heat Content`,
    `Average Sulfur Content`,
    `Average Ash Content`,
    `Fuel Cost`,
    `Moisture Content`,
    `Primary Transportation Mode`,
    `Year`,
    `Month`,
    `Plant ID`,
    `Operator ID`,
    `Fuel Cost`
  )

# Displaying modified dataframe
head(Page5)
```

EnergySource <chr>	Fuel Group <chr>	Quantity <dbl>	Average Heat Content <dbl>	Average Sulfur Content <dbl>	Ave
BIT	Coal	65880	20.310	0.75	
NG	Natural Gas	679779	1.023	0.00	
NG	Natural Gas	6759960	1.023	0.00	

EnergySource <chr>	Fuel Group <chr>	Quantity <dbl>	Average Heat Content <dbl>	Average Sulfur Content <dbl>	Ave
NG	Natural Gas	544609	1.035	0.00	
BIT	Coal	63686	24.442	1.21	
BIT	Coal	11611	24.376	2.21	

6 rows | 1-6 of 13 columns

Hide

```
# Removing December

Page1_New <- Page1 %>%
  filter(Month != 12 & !is.na('Reported Fuel Type Code'))

Page5_New <- Page5 %>%
  filter(Month != 12 & !is.na('Energy Source'))
```

Hide

```
#Confirming both months are stored as numbers
unique(Page1$Month)
```

```
 [1]  1  2  3  4  5  6  7  8  9 10 11 12
```

Hide

```
unique(Page5$Month)
```

```
 [1]  1  2  3  4  5  6  7  8  9 10 11
```

Hide

```
# Selecting relevant columns
Page1_New <- Page1 %>%
  select(`Month`, `Year`, `Plant ID`, `Operator ID`,
         `Quantity Consumed in Physical Units for Electric Generation`,
         `Physical Unit Label`, `Reported Prime Mover`, `ReportedFuelTypeCode`)

Page5_New <- Page5 %>%
  select(`Month`, `Year`, `Plant ID`, `Operator ID`, `EnergySource`, `Fuel Group`,
         Quantity, `Average Heat Content`, `Average Sulfur Content`,
         `Average Ash Content`, `Fuel Cost`, `Moisture Content`,
         `Primary Transportation Mode`)



# Merging
Merged_Data <- left_join(Page1_New, Page5_New,
                         by = c("Plant ID", "Operator ID", "Month", "Year")) %>%
  filter(ReportedFuelTypeCode == EnergySource)
```

Warning: Detected an unexpected many-to-many relationship between `x` and `y`.

Hide

```
# New merged data
head(Merged_Data)
```

M... <dbl>	Y... <dbl>	Plant ID <dbl>	Operator ID <dbl>	Quantity Consumed in Physical Units for Electric Generation <chr>
1	2024	3	195	41646.0
1	2024	3	195	41646.0
2	2024	3	195	19428.0
2	2024	3	195	19428.0
3	2024	3	195	150437.0
3	2024	3	195	150437.0

6 rows | 1-5 of 17 columns

Hide

```
write.csv(Merged_Data, "Merged_Dataset.csv", row.names = FALSE)
```