FALL 2018

TEXAS
The University of Texas at Austin

APPLIED ENGINEERING DATA ANALYSIS, OPTIMIZATION AND VISUALIZATION

EIA data v2: Let's value add this data

JOSHUA RHODES, PHD
Research Fellow/Adjunct Professor, The University of Texas at Austin

In EIA data v1 we did a lot

- Read in two different power plant datasets
- Aggregated data in both sets up to common attributes
  - Utility ID/Operator Id
  - Plant Code/Plant Id
  - Utility Name/Operator Name
  - Energy Source 1/Reported Fuel Type Code
- Merged the two datasets
- Calculated capacity factors
- Did sanity checks along the way
  - VERY IMPORTANT!
  - You must know your data

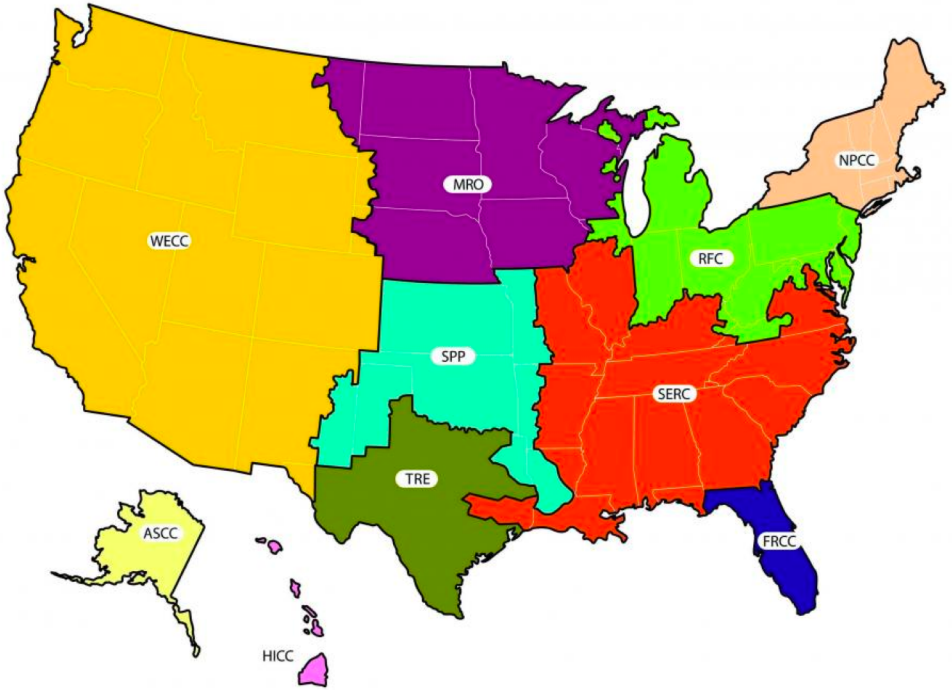
In this lecture we will start to add some more value to the dataset

- We lost the power plant names (nice to have)
- We would like the power plant location and NERC region
  - Latitude/Longitude
  - State

NERC regions are different sections of the US grid that help coordinate resources and balance power flows

Census.Region	NERC.Region
ESC	SERC
ESC	SERC
ESC	SERC
ESC	SERC
ESC	SERC
ESC	SERC

:ode Reserved.1 Reserved.2



It turns out that there is another dataset that has a bunch of that info. Yay!

- EIA 860M data
  - https://www.eia.gov/electricity/data/eia860m/
  - Here using December 2016 data
    - Use December values for each of the other years ;)

Generally the same concept

- Get rid of extra rows/columns
- Save as CSV
- Looks like:
- Read into R:
  - Entity.ID?!?!?! :/ sigh…

	A	B	C	D	E	F	G	H
1	Entity ID	Entity Name	Plant ID	Plant Name	Sector	Plant State	Generator ID	Nameplate C
2	195	Alabama Pov	2	Bankhead Da	Electric Utilit	AL	1	53.9
3	195	Alabama Pov	3	Barry	Electric Utilit	AL	1	153.1
4	195	Alabama Pov	3	Barry	Electric Utilit	AL	2	153.1
5	195	Alabama Pov	3	Barry	Electric Utilit	AL	4	403.7
6	195	Alabama Pov	3	Barry	Electric Utilit	AL	5	788.8
7	195	Alabama Pov	3	Barry	Electric Utilit	AL	A1CT	170.1
8	195	Alabama Pov	3	Barry	Electric Utilit	AL	A1CT2	170.1
9	195	Alabama Pov	3	Barry	Electric Utilit	AL	A1ST	195.2

```
> f860m <- read.csv('F860M.csv')
> head(f860m)
  Entity.ID      Entity.Name Plant.ID   Plant.Name
1       195 Alabama Power Co        2 Bankhead Dam
2       195 Alabama Power Co        3        Barry
3       195 Alabama Power Co        3        Barry
4       195 Alabama Power Co        3        Barry
5       195 Alabama Power Co        3        Barry
6       195 Alabama Power Co        3        Barry
```

Let's get what we want from the F860M data

```
> f860m2 <- f860m[c('Entity.ID', 'Plant.ID', 'Prime.Mover.Code', 'Latitude', 'Longitude', 'County', 'Balancing.Authority.Code')]
> head(f860m2)
  Entity.ID Plant.ID Prime.Mover.Code Latitude Longitude     County Balancing.Authority.Code
1       195        2               HY 33.45867 -87.35682 Tuscaloosa                     SOCO
2       195        3               ST 31.00690 -88.01030     Mobile                     SOCO
3       195        3               ST 31.00690 -88.01030     Mobile                     SOCO
4       195        3               ST 31.00690 -88.01030     Mobile                     SOCO
5       195        3               ST 31.00690 -88.01030     Mobile                     SOCO
6       195        3               CT 31.00690 -88.01030     Mobile                     SOCO
>
```

```
> dim(f860m2)
[1] 20754     7
> dim(m3)
[1] 8756   11
```

KEEP CHECKING THESE!

In this case we don't have to aggregate, just get rid of duplicates

```
> f860m3 <- f860m2[!duplicated(f860m2),]
> dim(f860m3)
[1] 9146    7
> head(f860m3)
   Entity.ID Plant.ID Prime.Mover.Code Latitude Longitude     County Balancing.Authority.Code
1        195        2               HY 33.45867 -87.35682 Tuscaloosa                     SOCO
2        195        3               ST 31.00690 -88.01030     Mobile                     SOCO
6        195        3               CT 31.00690 -88.01030     Mobile                     SOCO
8        195        3               CA 31.00690 -88.01030     Mobile                     SOCO
12       195        4               HY 32.58389 -86.28306     Elmore                     SOCO
15       195        7               ST 34.01280 -85.97080     Etowah                     SOCO
>
```

A straight merge is duplicating rows… that is not good…

```
> m4 <- merge(y = f860m3, x = m3, by.y = c('Entity.ID', 'Plant.ID'), by.x = c('Utility.ID', 'Plant.Code'))
> dim(m4)
[1] 11151    16
> summary(m4)
   Utility.ID      Plant.Code     Prime.Mover   Energy.Source.1                                Technology   Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min
 Min.   :    8   Min.   :    3   IC     :1641   NG     :4129   Natural Gas Fired Combined Cycle    :2217   Min.   :   0.1     Min.   :1896       Min.   :1896       Min.   :1891
 1st Qu.:10071   1st Qu.: 3482   PV     :1566   SUN    :1581   Solar Photovoltaic                  :1566   1st Qu.:   4.6     1st Qu.:1978       1st Qu.:1982       1st Qu.:1975
 Median :17650   Median :50933   GT     :1554   WAT    :1503   Conventional Hydroelectric          :1451   Median :  23.8     Median :2000       Median :2001       Median :1999
 Mean   :28342   Mean   :32987   HY     :1451   DFO    :1312   Petroleum Liquids                   :1420   Mean   : 154.4     Mean   :1991       Mean   :1993       Mean   :1990
 3rd Qu.:56516   3rd Qu.:57116   ST     :1421   WND    : 972   Natural Gas Fired Combustion Turbine:1143   3rd Qu.: 155.0     3rd Qu.:2010       3rd Qu.:2011       3rd Qu.:2010
 Max.   :60581   Max.   :60965   CA     :1133   LFG    : 394   Onshore Wind Turbine                : 971   Max.   :6495.0     Max.   :2016       Max.   :2016       Max.   :2016
                                 (Other):2385   (Other):1260   (Other)                             :2383                      NA's   :1          NA's   :1          NA's   :1
 Net.Generation..Megawatthours.     cap_fac         Prime.Mover.Code    Latitude       Longitude                 County      Balancing.Authority.Code
 Min.   : -768620              Min.   :-1.83357   IC     :1625     Min.   :18.97   Min.   :-171.71   Los Angeles   : 281   MISO   :1920
 1st Qu.:    2844              1st Qu.: 0.03214   PV     :1566     1st Qu.:34.54   1st Qu.:-111.30   Kern          : 144   PJM    :1517
 Median :   27419              Median : 0.22523   GT     :1559     Median :39.07   Median : -91.71   San Bernardino: 130   CISO   :1211
 Mean   :  519396              Mean   : 0.27697   HY     :1454     Mean   :38.55   Mean   : -95.17   Maricopa      : 124   ISNE   : 834
 3rd Qu.:  263928              3rd Qu.: 0.44224   ST     :1408     3rd Qu.:42.20   3rd Qu.: -80.09   Middlesex     : 114   SWPP   : 826
 Max.   :32377477              Max.   : 1.85495   CA     :1142     Max.   :71.29   Max.   : -40.47   Jefferson     : 112   (Other):4838
                                                  (Other):2397     NA's   :7       NA's   :7         (Other)       :10246   NA's   :   5

> sum(m4$Nameplate.Capacity)
[1] 1721918

> sum(m3$Nameplate.Capacity)
[1] 1153254
>
```

Hmmmm :/

We don't need prime mover here, so lets get rid of and remove duplicates again, merge, & check

```
> f860m4 <- subset(f860m3, select=-c(Prime.Mover.Code))
> head(f860m4)
   Entity.ID Plant.ID Latitude Longitude     County Balancing.Authority.Code
1        195        2 33.45867 -87.35682 Tuscaloosa                     SOCO
2        195        3 31.00690 -88.01030     Mobile                     SOCO
6        195        3 31.00690 -88.01030     Mobile                     SOCO
8        195        3 31.00690 -88.01030     Mobile                     SOCO
12       195        4 32.58389 -86.28306     Elmore                     SOCO
15       195        7 34.01280 -85.97080     Etowah                     SOCO
```

```
> f860m5 <- f860m4[!duplicated(f860m4),]
```

```
> m4 <- merge(x = m3, y = f860m5, by.x = c('Utility.ID', 'Plant.Code'), by.y = c('Entity.ID', 'Plant.ID'), all.x = F, all.y = F)
> summary(m4)
   Utility.ID      Plant.Code     Prime.Mover   Energy.Source.1                                  Technology   Nameplate.Capacity Operc
 Min.   :    8   Min.   :    3   PV     :1529   NG     :2328   Solar Photovoltaic                    :1529   Min.   :   0.1     Min.
 1st Qu.:11208   1st Qu.: 4075   HY     :1396   SUN    :1544   Conventional Hydroelectric            :1396   1st Qu.:   3.6     1st (
 Median :19436   Median :54632   IC     :1352   WAT    :1435   Petroleum Liquids                     :1043   Median :  16.0     Medic
 Mean   :30691   Mean   :35331   GT     :1095   DFO    : 978   Natural Gas Fired Combined Cycle      : 988   Mean   : 133.6     Mean
 3rd Qu.:56996   3rd Qu.:57734   ST     :1036   WND    : 950   Onshore Wind Turbine                  : 949   3rd Qu.: 102.4     3rd (
 Max.   :60581   Max.   :60965   WT     : 949   LFG    : 346   Natural Gas Fired Combustion Turbine: 802   Max.   :6495.0     Max.
                                 (Other):1213   (Other): 989   (Other)                               :1863                      NA's

 Net.Generation..Megawatthours.     cap_fac           Latitude       Longitude                County     Balancing.Authority.Code
 Min.   : -768620               Min.   :-1.83357   Min.   :18.97   Min.   :-171.71   Los Angeles   : 170   MISO   :1474
 1st Qu.:    2722               1st Qu.: 0.04415   1st Qu.:35.05   1st Qu.:-111.89   Kern          : 138   PJM    :1189
 Median :   19454               Median : 0.22743   Median :39.37   Median : -92.04   San Bernardino: 120   CISO   :1054
 Mean   :  452496               Mean   : 0.27649   Mean   :38.94   Mean   : -95.37   Jefferson     :  77   ISNE   : 650
 3rd Qu.:  177745               3rd Qu.: 0.42654   3rd Qu.:42.44   3rd Qu.: -79.92   Middlesex     :  77   SWPP   : 622
 Max.   :32377477               Max.   : 1.85495   Max.   :71.29   Max.   : -40.47   Riverside     :  71   (Other):3576
                                                   NA's   :7       NA's   :7         (Other)       :7917   NA's   :   5
```

```
> sum(m3$Nameplate.Capacity)
[1] 1153254
> sum(m4$Nameplate.Capacity)
[1] 1145062
>
```

Sometimes merges w/ incomplete information yield messy data, let's check those

```
> summary(m4)
   Utility.ID      Plant.Code     Prime.Mover   Energy.Source.1                                Technology  Nameplate.Capacity
 Min.   :    8   Min.   :    3   PV     :1529   NG     :2328   Solar Photovoltaic                  :1529  Min.   :   0.1
 1st Qu.:11208   1st Qu.: 4075   HY     :1396   SUN    :1544   Conventional Hydroelectric          :1396  1st Qu.:   3.6
 Median :19436   Median :54632   IC     :1352   WAT    :1435   Petroleum Liquids                   :1043  Median :  16.0
 Mean   :30691   Mean   :35331   GT     :1095   DFO    : 978   Natural Gas Fired Combined Cycle    : 988  Mean   : 133.6
 3rd Qu.:56996   3rd Qu.:57734   ST     :1036   WND    : 950   Onshore Wind Turbine                : 949  3rd Qu.: 102.4
 Max.   :60581   Max.   :60965   WT     : 949   LFG    : 346   Natural Gas Fired Combustion Turbine: 802  Max.   :6495.0
                                 (Other):1213   (Other): 989   (Other)                             :1863
 Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.     cap_fac          Latitude
 Min.   :1896       Min.   :1896       Min.   :1891       Min.   : -768620               Min.   :-1.83357   Min.   :18.97
 1st Qu.:1978       1st Qu.:1983       1st Qu.:1975       1st Qu.:    2722               1st Qu.: 0.04415   1st Qu.:35.05
 Median :2001       Median :2002       Median :2001       Median :   19454               Median : 0.22743   Median :39.37
 Mean   :1992       Mean   :1993       Mean   :1990       Mean   :  452496               Mean   : 0.27649   Mean   :38.94
 3rd Qu.:2011       3rd Qu.:2012       3rd Qu.:2011       3rd Qu.:  177745               3rd Qu.: 0.42654   3rd Qu.:42.44
 Max.   :2016       Max.   :2016       Max.   :2016       Max.   :32377477               Max.   : 1.85495   Max.   :71.29
 NA's   :1          NA's   :1          NA's   :1                                                             NA's   :7
   Longitude                 County     Balancing.Authority.Code
 Min.   :-171.71   Los Angeles   : 170   MISO   :1474
 1st Qu.:-111.89   Kern          : 138   PJM    :1189
 Median : -92.04   San Bernardino: 120   CISO   :1054
 Mean   : -95.37   Jefferson     :  77   ISNE   : 650
 3rd Qu.: -79.92   Middlesex     :  77   SWPP   : 622
 Max.   : -40.47   Riverside     :  71   (Other):3576
 NA's   :7         (Other)       :7917   NA's   :   5
```

```
> m4[!complete.cases(m4),]
     Utility.ID Plant.Code Prime.Mover Energy.Source.1          Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max
87        10433      60250          IC             DFO   Petroleum Liquids               8.4           2006.333                2015
2237       1857       6567          IC             DFO   Petroleum Liquids               6.4           2005.250                2012
2994        221      60260          IC             DFO   Petroleum Liquids               1.1           2010.000                2010
4896      56226      60696          IC             LFG         Landfill Gas               3.4           2008.000                2008
5671      57170      60656          BA             MWH           Batteries               19.8           2016.000                2016
6300      58500      58525          IC             OBG  Other Waste Biomass               1.1                 NA                  NA
7433      60005      58735          PV             SUN   Solar Photovoltaic               4.0           2015.000                2015
7574      60336      60567          BA             MWH           Batteries               2.0           2015.000                2015
7653      60430      60700          WT             WND Onshore Wind Turbine               5.0           2015.000                2015
7719      60529      60878          PV             SUN   Solar Photovoltaic               1.9           2016.000                2016
7720      60530      60879          PV             SUN   Solar Photovoltaic               2.0           2015.000                2015
8118       7601      60874          PV             SUN   Solar Photovoltaic               5.0           2016.000                2016
8120       7601      60877          PV             SUN   Solar Photovoltaic               5.0           2016.000                2016
     Operating.Year.min Net.Generation..Megawatthours.      cap_fac Latitude  Longitude          County Balancing.Authority.Code
87                 2002                          0.000 0.0000000000 57.77556 -152.48028    Kodiak Island                    <NA>
2237               2001                      12560.000 0.2240296804 41.17556  -71.57110       Washington                    <NA>
2994               2010                       1137.000 0.1179950187 65.33172 -166.47951             Nome                    <NA>
4896               2008                        855.000 0.0287066882       NA         NA                                        
5671               2016                     149060.000 0.8593930169       NA         NA          McHenry                     PJM
6300                 NA                       7830.517 0.8126314861 37.63353 -122.13972          Alameda                    CISO
7433               2015                       7022.000 0.2003995434 35.54194  -80.36333           Rowan                    <NA>
7574               2015                         -2.000 -0.0001141553       NA         NA San Diego County                    CISO
7653               2015                      12742.000 0.2909132420 41.96107  -73.14078                                    <NA>
7719               2016                        162.000 0.0097332372       NA         NA       Worcester                    ISNE
7720               2015                       2516.000 0.1436073059       NA         NA       Berkshire                    ISNE
8118               2016                         37.000 0.0008447489       NA         NA         Windsor                    ISNE
8120               2016                          8.000 0.0001826484       NA         NA                                    ISNE
>
```

That is pretty good. You will need to make these checks/decisions about your data

- Add year
- This is pretty good, now we can explore some

```
> data <- m4[complete.cases(m4),]
> data$year <- 2016
> summary(data)
   Utility.ID      Plant.Code     Prime.Mover   Energy.Source.1                                  Technology   Nameplate.Capacity
 Min.   :    8   Min.   :    3   PV     :1524   NG     :2328   Solar Photovoltaic                     :1524   Min.   :   0.1
 1st Qu.:11208   1st Qu.: 4072   HY     :1396   SUN    :1539   Conventional Hydroelectric             :1396   1st Qu.:   3.6
 Median :19436   Median :54624   IC     :1347   WAT    :1435   Petroleum Liquids                      :1040   Median :  16.0
 Mean   :30679   Mean   :35299   GT     :1095   DFO    : 975   Natural Gas Fired Combined Cycle       : 988   Mean   : 133.8
 3rd Qu.:56995   3rd Qu.:57723   ST     :1036   WND    : 949   Onshore Wind Turbine                   : 948   3rd Qu.: 102.5
 Max.   :60581   Max.   :60965   WT     : 948   LFG    : 345   Natural Gas Fired Combustion Turbine: 802   Max.   :6495.0
                                 (Other):1211   (Other): 986   (Other)                                :1859
 Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.     cap_fac          Latitude
 Min.   :1896       Min.   :1896       Min.   :1891       Min.   : -768620              Min.   :-1.8336   Min.   :18.97
 1st Qu.:1978       1st Qu.:1983       1st Qu.:1974       1st Qu.:    2729              1st Qu.: 0.0444   1st Qu.:35.05
 Median :2001       Median :2002       Median :2000       Median :   19558              Median : 0.2276   Median :39.37
 Mean   :1991       Mean   :1993       Mean   :1990       Mean   :  453160              Mean   : 0.2766   Mean   :38.93
 3rd Qu.:2011       3rd Qu.:2012       3rd Qu.:2011       3rd Qu.:  178852              3rd Qu.: 0.4267   3rd Qu.:42.44
 Max.   :2016       Max.   :2016       Max.   :2016       Max.   :32377477              Max.   : 1.8549   Max.   :71.29

   Longitude                 County     Balancing.Authority.Code      year
 Min.   :-171.71   Los Angeles    : 170   MISO   :1474             Min.   :2016
 1st Qu.:-111.89   Kern           : 138   PJM    :1188             1st Qu.:2016
 Median : -92.04   San Bernardino: 120   CISO   :1052             Median :2016
 Mean   : -95.36   Jefferson      :  77   ISNE   : 646             Mean   :2016
 3rd Qu.: -79.92   Middlesex      :  77   SWPP   : 622             3rd Qu.:2016
 Max.   : -40.47   Riverside      :  71   NYIS   : 406             Max.   :2016
                   (Other)        :7904   (Other):3169
```

Now we can explore the data

```
> sum(data$Net.Generation..Megawatthours.)
[1] 3877693734
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Natural Gas Fired Combined Cycle'])
[1] 1141726068
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Solar Photovoltaic'])
[1] 31675088
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Onshore Wind Turbine'])
[1] 224683699
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Batteries'])
[1] -157340
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal'])
[1] 1122423660
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal'])/sum(data$Net.Generation..Megawatthours.)
[1] 0.2894565
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Natural Gas Fired Combined Cycle'])/sum(data$Net.Generation..Megawatthours.)
[1] 0.2944343
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Onshore Wind Turbine'])/sum(data$Net.Generation..Megawatthours.)
[1] 0.05794261
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Solar Photovoltaic'])/sum(data$Net.Generation..Megawatthours.)
[1] 0.008168538
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Nuclear'])/sum(data$Net.Generation..Megawatthours.)
[1] 0.2065183
>
```

We can explore the data on multiple levels

```
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'PJM'])
[1] 255671591
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO'])
[1] 64990688
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'CISO'])
[1] 317470.5
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Solar Photovoltaic' & data$Balancing.Authority.Code == 'CISO'])
[1] 17217228
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Solar Photovoltaic' & data$Balancing.Authority.Code == 'ERCO'])
[1] 698305
>
```

```
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.avg < 1970])
[1] 0
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.avg < 1975])
[1] 0
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.avg < 1980])
[1] 22264310
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.avg < 1978])
[1] 2423894
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.avg < 1977])
[1] 0
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.min < 1978])
[1] 29453945
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.min < 1970])
[1] 0
> sum(data$Net.Generation..Megawatthours.[data$Technology == 'Conventional Steam Coal' & data$Balancing.Authority.Code == 'ERCO' &
data$Operating.Year.min < 1975])
[1] 7189635
>
```

Write that file

```
> write.csv(data, 'complete_eia_data_2016.csv', row.names = F)
>
```

Now we can start looking at data w/ a spatial dimension

```
> unique(data$Technology[data$County == 'Los Angeles'])
 [1] Natural Gas Fired Combustion Turbine   Conventional Hydroelectric            Hydroelectric Pumped Storage
 [4] Natural Gas Fired Combined Cycle       Natural Gas Steam Turbine             Solar Photovoltaic
 [7] Other Waste Biomass                    Landfill Gas                          Natural Gas Internal Combustion Engine
[10] Batteries                              Petroleum Liquids                     All Other
[13] Other Gases                            Other Natural Gas
26 Levels: All Other Batteries Coal Integrated Gasification Combined Cycle ... Wood/Wood Waste Biomass
> unique(data$Technology[data$County == 'Travis'])
[1] Natural Gas Fired Combustion Turbine Natural Gas Steam Turbine           Petroleum Liquids
[4] Natural Gas Fired Combined Cycle     Conventional Hydroelectric          Landfill Gas
[7] Solar Photovoltaic
26 Levels: All Other Batteries Coal Integrated Gasification Combined Cycle ... Wood/Wood Waste Biomass
>
```

Other aspects we didn't get into the data

- Operating status <- could be interesting
- Planned retirement year/month
- Plant name – lost along the way
- State…

Next assignment: replicate this ++

- You will essentially replicate the workflow in these *two* lectures to recreate the final product
  - Add Plant Name back in there
  - Add State in
  - Deliverables will be code that writes a file and gives summary
  - See Canvas for more information on assignment
- Hint: make your code reproducible, we will be doing other years next…