FALL 2018

APPLIED ENGINEERING DATA ANALYSIS, OPTIMIZATION AND VISUALIZATION

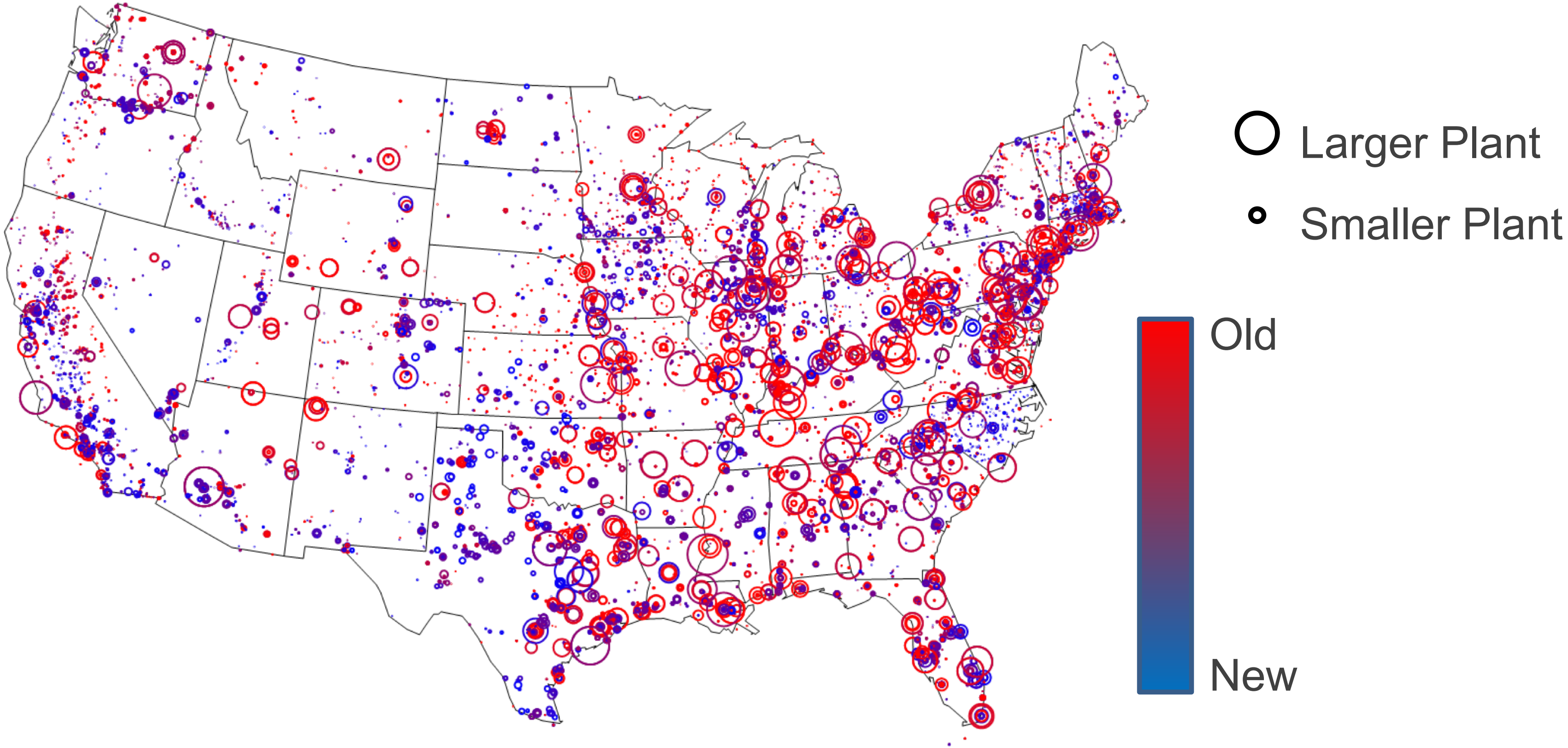
EIA data, let's get our hands dirty…

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

Now we will get more familiar with data and R using EIA's power plant database

- There are two sources of data we will clean and merge
- EIA Form 860: https://www.eia.gov/electricity/data/eia860/
- EIA Form 923: https://www.eia.gov/electricity/data/eia923/
- Each have different information about the same power plants, but they are not perfectly aligned :/

Merged data can let you build maps like this



Goal calculate the capacity factor of each power plant in the EIA database

- Capacity factor (%) = MWh(yr_n)/(8760*nam_cap)
  - MWh(yr_n) (MWh) = amount of energy produced by each plant in year n (found in F923 data)
  - nam_cap (MW) = name plate capacity of plant (found in F860 data)
  - 8760 = number of hours in a year
  - Capacity factor = %

Some complications…

- For 2016 data (will be different for other years, yay!)
  - F923 has data by power plant (n = 12,559)
  - F860 has data by generator (n = 20,694)
  - Some plants have > 1 generator
  - This could get tricky

Let's get the right data (my process for 2016)

- In the zipped folders, you will find multiple files, get:
  - From 860 data get: 3_1_Generator_Y2016.xlsx file
    - Tab: Operable
    - Here renamed F860.csv
  - From 923 data get: EIA923_Schedules_2_3_4_5_M_12_2016.xlsx
    - Tab: Generation and Fuel Data
    - Here renamed F923.csv

Now we run into our first problem

- Different column names
  - Utility ID (F860) = Operator Id (F923) ??
  - Plant Code (F860) = Plant Id (F923) ??
  - Utility Name (F860) = Operator Name (F923) ??
  - Energy Source 1 (F860) = Reported Fuel Type Code (F923) ??

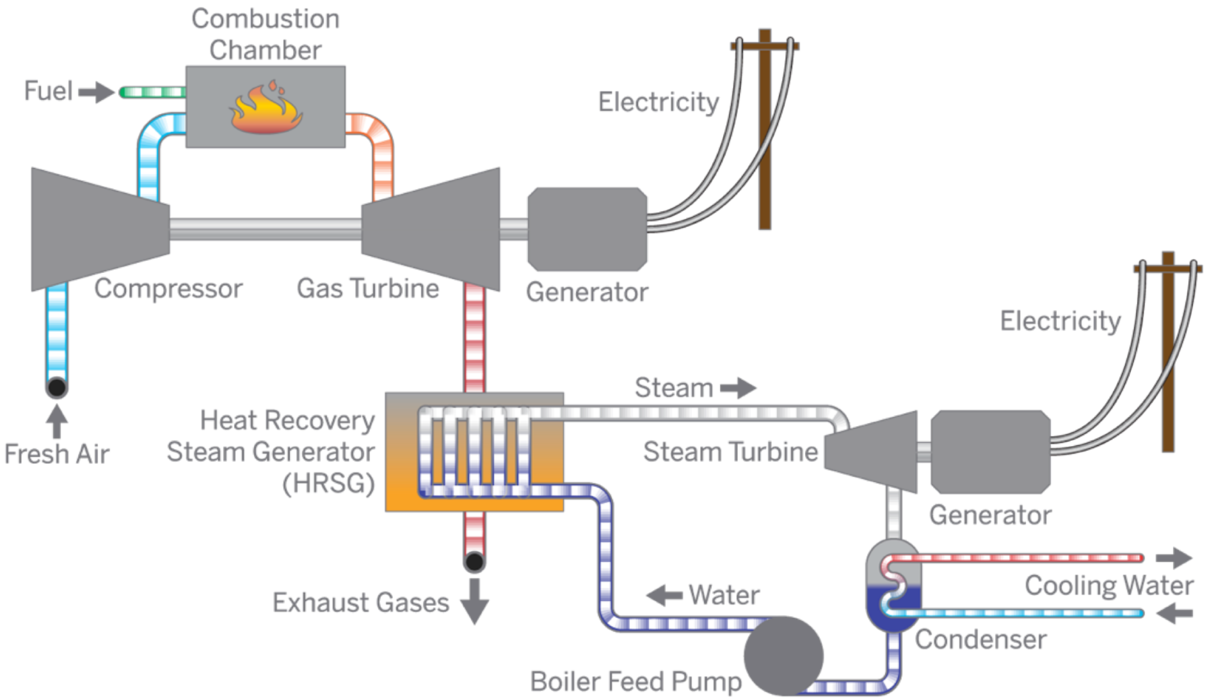
Also different number of items in each

- F860: n = 20,694
  - (2016 only!! will be diff for ’17, ’15, ’14, etc.)
- F923: n = 12,559
  - Same caveat as above

This is where knowing your data is key

- For: EIA923_Schedules_2_3_4_5_M_12….xlsx
  - Tab: Page 7 File Layout
    - Reported Primer Mover
      - CA = Combined-Cycle -- Steam Part
      - CT = Combined-Cycle Combustion Turbine Part
      - Etc.
    - Reported Fuel Type
      - NG = Natural Gas
      - BIT = Bituminous coal
      - Etc.
  - “Natural Gas Fired Combined Cycle” (F860 “Technology”) has both a CA and CT part

And that is just exactly what a combined cycle is



1/ Some more figuring out, let's take a look at the plant named: Barry

- Some power plants have multiple of a type
- F860: 10 rows of data

Utility ID	Utility Name	Plant Code	Plant Name	State	County	Generator ID	Technology	Prime Mover	Unit Code	Ownership	Duct Burners	Can Bypass H	RTO/ISO LMI	RTO/ISO Loca	Nameplate C	N
195	Alabama Power Co	2	Bankhead Dam	AL	Tuscaloosa	1	Conventional Hydroelectric	HY		S	X	X			53.9	
195	Alabama Power Co	3	Barry	AL	Mobile	1	Natural Gas Steam Turbine	ST		S	X	X			153.1	
195	Alabama Power Co	3	Barry	AL	Mobile	2	Natural Gas Steam Turbine	ST		S	X	X			153.1	
195	Alabama Power Co	3	Barry	AL	Mobile	4	Conventional Steam Coal	ST		S	X	X			403.7	
195	Alabama Power Co	3	Barry	AL	Mobile	5	Conventional Steam Coal	ST		S	X	X			788.8	
195	Alabama Power Co	3	Barry	AL	Mobile	A1CT	Natural Gas Fired Combined Cycle	CT	G521	S	X	N			170.1	
195	Alabama Power Co	3	Barry	AL	Mobile	A1CT2	Natural Gas Fired Combined Cycle	CT	G521	S	X	N			170.1	
195	Alabama Power Co	3	Barry	AL	Mobile	A1ST	Natural Gas Fired Combined Cycle	CA	G521	S	Y	X			195.2	
195	Alabama Power Co	3	Barry	AL	Mobile	A2C1	Natural Gas Fired Combined Cycle	CT	G522	S	X	N			170.1	
195	Alabama Power Co	3	Barry	AL	Mobile	A2C2	Natural Gas Fired Combined Cycle	CT	G522	S	X	N			170.1	
195	Alabama Power Co	3	Barry	AL	Mobile	A2ST	Natural Gas Fired Combined Cycle	CA	G522	S	Y	X			195.2	
195	Alabama Power Co	4	Walter Bouldin Dam	AL	Elmore	1	Conventional Hydroelectric	HY		S	X	X			75	

2/ Some more figuring out, let’s take a look at the plant named: Barry

- On the F923 side: 4 rows of data

1	Plant Id	Combined Heat And Power Plant	Nuclear Unit	Plant Name	Operator Name	Operator Id	Plant State	Census Regic	NERC Region	Reserved	NAICS Code	EIA Sector Nı	Sector Name	Reported Prime Mover	Reported Fuel Type Code
2	3	N		Barry	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utilit	CA	NG
3	3	N		Barry	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utilit	CT	NG
4	3	N		Barry	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utilit	ST	BIT
5	3	N		Barry	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utilit	ST	NG
6	4	N		Walter Bouldin Dam	Alabama Power Co	195	AL	ESC	SERC		22	1	Electric Utilit	HY	WAT

The logic goes like this

- We have the following to reconcile:
  - Utility ID / Operator Id
  - Plant Code / Plant Id
  - Utility Name / Operator Name
  - Energy Source 1 / Reported Fuel Type Code
  - Prime Mover (PM)

Remember we are trying to find the capacity factor for each power plant / technology

- Barry has 3 power plants types on site (F860)
  - Natural gas (NG)
    - Natural Gas Fired Combined Cycle (PM: CA & CT)
    - Natural Gas Steam Turbine (PM: ST)
  - Coal (BIT)
    - Conventional Steam Coal (PM: ST)

More Barry

- It looks like (from F860:Generator ID)
  - There are two CT units for every CA unit
  - Not always the case
    - F860:Plant Name:Kyrene = 1 CT and 1 CA
- Unfortunately, F923 does not have a Generator ID value, it aggregates by Fuel Type and Prime Mover
  - Thus, we will be limited to power plant types per Plant ID
    - Barry (10 total generators -> 3 aggregated generators)
      - Natural Gas Fired Combined Cycle (looks like 2 units (4 CT & 2 CA total), but we have to aggregate to 1)
      - Natural Gas Steam Turbine (looks like 2 turbines, but we aggregate to 1)
      - Conventional Steam Coal (looks like 2 different size units, but have to aggregate to 1)

When faced with this, we have to make a decision

- Since we want to merge and the number of columns are not the same, we can either
  - Add data to the lower count dataset (F923)
    - Hard to think when this is a good idea (unless you know data are missing)
  - Reduce count of higher count dataset (F860)
    - Here it makes sense to aggregate data (will reduce # of rows)

So here we go

```
> f860 <- read.csv("F860.csv")
> f860[1:5,1:5]
  Utility.ID      Utility.Name Plant.Code    Plant.Name State
1        195 Alabama Power Co          2 Bankhead Dam    AL
2        195 Alabama Power Co          3        Barry    AL
3        195 Alabama Power Co          3        Barry    AL
4        195 Alabama Power Co          3        Barry    AL
5        195 Alabama Power Co          3        Barry    AL
> f923 <- read.csv("F932.csv")
> f923[1:5,1:5]
  Plant.Id Combined.Heat.And.Power.Plant Nuclear.Unit.Id          Plant.Name    Operator.Name
1        3                             N              NA               Barry Alabama Power Co
2        3                             N              NA               Barry Alabama Power Co
3        3                             N              NA               Barry Alabama Power Co
4        3                             N              NA               Barry Alabama Power Co
5        4                             N              NA Walter Bouldin Dam Alabama Power Co
> |
```

Let's first do some data exploring

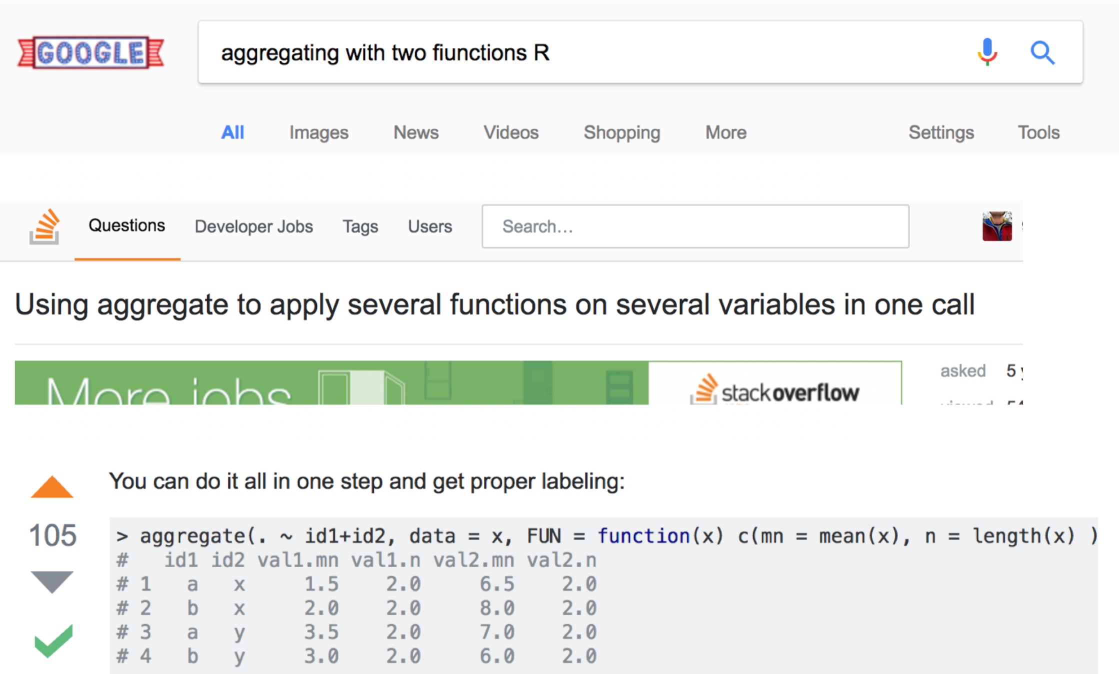
```
> names(f860)
 [1] "Utility.ID"
 [2] "Utility.Name"
 [3] "Plant.Code"
 [4] "Plant.Name"
 [5] "State"
 [6] "County"
 [7] "Generator.ID"
 [8] "Technology"
 [9] "Prime.Mover"
[10] "Unit.Code"
[11] "Ownership"
[12] "Duct.Burners"
[13] "Can.Bypass.Heat.Recovery.Steam.Generator."
[14] "RTO.ISO.LMP.Node.Designation"
[15] "RTO.ISO.Location.Designation.for.Reporting.Wholesale.Sales.Data.to.FERC"
[16] "Nameplate.Capacity..MW."
[17] "Nameplate.Power.Factor"
[18] "Summer.Capacity..MW."
[19] "Winter.Capacity..MW."
```

```
> sum(f860$Nameplate.Capacity..MW.)
[1] 1178148
> |
```

- There looks to be about 1.2 TW of power plants in the database
- That sounds about right
- Note: R fills non-alpha characters with ., i.e: " " -> "."

We know that NGCC plants have two types of prime movers, do others?

```
> ex1 <- as.data.frame(aggregate(x = f860$Nameplate.Capacity..MW., by = list(f860$Technology, f860$Prime.Mover), FUN = function(x) c(sum = sum(x), n = length(x))))
> ex1 <- ex1[order(ex1$Group.1),]
> head(ex1, n = 25)
                                   Group.1 Group.2    x.sum      x.n
22                               All Other      GT     39.1      1.0
34                               All Other      OT     90.9     13.0
41                               All Other      ST   1464.8     69.0
1                                Batteries      BA    561.7     67.0
3   Coal Integrated Gasification Combined Cycle CA  464.9      2.0
12  Coal Integrated Gasification Combined Cycle CT  665.9      3.0
28                 Conventional Hydroelectric   HY  79332.5   4046.0
42                    Conventional Steam Coal   ST 289398.7    844.0
21                               Flywheels      FW     42.0      3.0
2                               Geothermal      BT    861.0     91.0
43                              Geothermal      ST   2943.6    104.0
39               Hydroelectric Pumped Storage   PS  21648.7    156.0
4                             Landfill Gas      CA     65.3      8.0
13                            Landfill Gas      CT    132.5     23.0
18                            Landfill Gas      FC     11.5      7.0
23                            Landfill Gas      GT    325.5     80.0
29                            Landfill Gas      IC   1473.2   1508.0
44                            Landfill Gas      ST    189.5      8.0
45                    Municipal Solid Waste     ST   2701.2     91.0
5             Natural Gas Fired Combined Cycle  CA  98553.8    661.0
11            Natural Gas Fired Combined Cycle  CS  11909.4     55.0
14            Natural Gas Fired Combined Cycle  CT 162432.0   1160.0
24        Natural Gas Fired Combustion Turbine  GT 151915.3   2280.0
30       Natural Gas Internal Combustion Engine IC   4150.9   1033.0
46                  Natural Gas Steam Turbine   ST  83778.6    587.0
> |
```



Looks like NGCC plants come in 3 flavors

- We need to figure this out a bit more…
  - Do I want to lump all NGCC plants together?
  - What about landfill gas plants?
    - 6 Prime Movers there…

There are 27 different F860:Technology categories

```
> unique(ex1$Group.1)
 [1] All Other                                Batteries                                Coal Integrated Gasification Combined Cycle
 [4] Conventional Hydroelectric               Conventional Steam Coal                  Flywheels
 [7] Geothermal                               Hydroelectric Pumped Storage             Landfill Gas
[10] Municipal Solid Waste                    Natural Gas Fired Combined Cycle         Natural Gas Fired Combustion Turbine
[13] Natural Gas Internal Combustion Engine   Natural Gas Steam Turbine                Natural Gas with Compressed Air Storage
[16] Nuclear                                  Offshore Wind Turbine                    Onshore Wind Turbine
[19] Other Gases                              Other Natural Gas                        Other Waste Biomass
[22] Petroleum Coke                           Petroleum Liquids                        Solar Photovoltaic
[25] Solar Thermal with Energy Storage        Solar Thermal without Energy Storage     Wood/Wood Waste Biomass
27 Levels: All Other Batteries Coal Integrated Gasification Combined Cycle Conventional Hydroelectric Conventional Steam Coal Flywheels ... Wood/Wood Waste Biomass
> 
```

Since the F923 data are aggregated on Prime Mover & Plant Id & Fuel, we will do the same for F860 data

- Essentially what we want from the F860 data are:
  - Plant Code
  - Prime Mover
  - Technology
  - Nameplate Capacity (summed values)
- If we have this to merge w/ F923 data, we can then figure out capacity factors of plants

Let's start by reducing the data down

- We will do this a few times and merge back what we want (capacity, age, more?)

```
> f860_r1 <- f860[c('Utility.ID', 'Plant.Code', 'Technology', 'Prime.Mover', 'Energy.Source.1', 'Nameplate.Capacity..MW.')]
> head(f860_r1)
  Utility.ID Plant.Code                     Technology Prime.Mover Energy.Source.1 Nameplate.Capacity..MW.
1        195          2    Conventional Hydroelectric          HY             WAT                    53.9
2        195          3     Natural Gas Steam Turbine          ST              NG                   153.1
3        195          3     Natural Gas Steam Turbine          ST              NG                   153.1
4        195          3       Conventional Steam Coal          ST             BIT                   403.7
5        195          3       Conventional Steam Coal          ST             BIT                   788.8
6        195          3 Natural Gas Fired Combined Cycle       CT              NG                   170.1
> |
```

Doing so we reduce the F860 data from ~20k to ~9.4k

```
> ex2 <- as.data.frame(aggregate(x = f860_r1$Nameplate.Capacity..MW., by = list(f860_r1$Utility.ID, f860_r1$Plant.Code, f860_r1$Technology, f860_r1$Prime.Mover, f860_r1$Energy.Source.1), FUN = sum))
> names(ex2) <- c('Utility.ID', 'Plant.Code', 'Technology', 'Prime.Mover', 'Energy.Source.1', 'Nameplate.Capacity')
> head(ex2, n = 15)
   Utility.ID Plant.Code          Technology Prime.Mover Energy.Source.1 Nameplate.Capacity
1         135      10593 Other Waste Biomass          ST              AB               12.1
2       19946      50293 Other Waste Biomass          ST              AB               28.6
3       19365      50482 Other Waste Biomass          ST              AB               70.6
4       11415      51008 Other Waste Biomass          ST              AB                3.0
5       16098      54338 Other Waste Biomass          ST              AB               23.5
6       13492      54627 Other Waste Biomass          ST              AB              128.9
7         772      54908 Other Waste Biomass          ST              AB                9.8
8       56413      57119 Other Waste Biomass          ST              AB               23.4
9       58873      59035 Other Waste Biomass          ST              AB                9.3
10      56150      10475        Other Gases          OT             BFG               20.0
11      56147      10245        Other Gases          ST             BFG              176.2
12      56165      10397        Other Gases          ST             BFG              152.0
13       9454      10398        Other Gases          ST             BFG               68.0
14      17145      50067        Other Gases          ST             BFG               37.5
15      19519      50732        Other Gases          ST             BFG               52.5
>
```

```
> dim(ex2)
[1] 9392    6
> sum(ex2$Nameplate.Capacity)
[1] 1178148
>
```

Let’s do the same for Operating Year – we can assess the age of plants that way

```
> f860_r2 <- f860[c('Utility.ID', 'Plant.Code', 'Technology', 'Prime.Mover', 'Energy.Source.1', 'Operating.Year')]
> ex3 <- as.data.frame(aggregate(x = f860_r2$Operating.Year, by = list(f860_r2$Utility.ID, f860_r2$Plant.Code, f860_r2$Technology, f860_r2$Prime.Mover,
f860_r2$Energy.Source.1), FUN = function(x) c(avg = mean(x), max = max(x), min = min(x))))
> names(ex3)
[1] "Group.1" "Group.2" "Group.3" "Group.4" "Group.5" "x"
> summary(ex3$x)
      avg             max             min      
 Min.   :1896   Min.   :1896   Min.   :1891  
 1st Qu.:1980   1st Qu.:1984   1st Qu.:1976  
 Median :2001   Median :2002   Median :2001  
 Mean   :1992   Mean   :1994   Mean   :1990  
 3rd Qu.:2012   3rd Qu.:2012   3rd Qu.:2011  
 Max.   :2016   Max.   :2016   Max.   :2016  
 NA's   :1      NA's   :1      NA's   :1     
> summary(as.data.frame(ex3$x[,1]))
   ex3$x[, 1]  
 Min.   :1896  
 1st Qu.:1980  
 Median :2001  
 Mean   :1992  
 3rd Qu.:2012  
 Max.   :2016  
 NA's   :1     
> |
```

- Issue here with the way the aggregate function handles multiple operations
- Need to break apart the ’x’ into 3 columns: avg, max, min of Op. Year

There is likely a better way to do this, if you find one (in R), then let me know…

```
> ex4 <- as.data.frame(cbind(ex3[c('Group.1', 'Group.2', 'Group.3', 'Group.4', 'Group.5')], as.data.frame(as.matrix(ex3$x))))
> names(ex4) <- c('Utility.ID', 'Plant.Code', 'Technology', 'Prime.Mover', 'Energy.Source.1', 'Operating.Year.avg', 'Operating.Year.max', 'Operating.Year.min')
> dim(ex4)
[1] 9392    8
> head(ex4, n = 15)
   Utility.ID Plant.Code          Technology Prime.Mover Energy.Source.1 Operating.Year.avg Operating.Year.max Operating.Year.min
1         135      10593 Other Waste Biomass          ST              AB           1984.000               1984               1984
2       19946      50293 Other Waste Biomass          ST              AB           1989.000               1989               1989
3       19365      50482 Other Waste Biomass          ST              AB           2003.500               2007               1997
4       11415      51008 Other Waste Biomass          ST              AB           1981.000               1981               1981
5       16098      54338 Other Waste Biomass          ST              AB           1987.500               2009               1973
6       13492      54627 Other Waste Biomass          ST              AB           2001.000               2006               1996
7         772      54908 Other Waste Biomass          ST              AB           1983.000               1983               1983
8       56413      57119 Other Waste Biomass          ST              AB           2009.000               2009               2009
9       58873      59035 Other Waste Biomass          ST              AB           2016.000               2016               2016
10      56150      10475        Other Gases           OT             BFG           1981.000               1981               1981
11      56147      10245        Other Gases           ST             BFG           1969.333               1970               1969
12      56165      10397        Other Gases           ST             BFG           1962.800               2002               1939
13       9454      10398        Other Gases           ST             BFG           1971.000               2010               1950
14      17145      50067        Other Gases           ST             BFG           1985.000               1985               1985
15      19519      50732        Other Gases           ST             BFG           1962.667               2002               1943
> 
```

Now we merge the two sets of F860 data back together

```
> m1 <- merge(x = ex2, y = ex4, by = c('Utility.ID', 'Plant.Code', 'Technology', 'Prime.Mover', 'Energy.Source.1'))
> dim(m1)
[1] 9392    9
> head(m1)
  Utility.ID Plant.Code                      Technology Prime.Mover Energy.Source.1 Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min
1      10000       1241            Conventional Steam Coal          ST             SUB             1578.0               1975               1977               1973
2      10000       2079            Conventional Steam Coal          ST             SUB              594.3               1969               1969               1969
3      10000       2079     Natural Gas Fired Combined Cycle          CA              NG              146.0               2000               2000               2000
4      10000       2079     Natural Gas Fired Combined Cycle          CT              NG              176.0               1997               1997               1997
5      10000       2079 Natural Gas Fired Combustion Turbine          GT              NG              154.0               2000               2000               2000
6      10000       2080            Conventional Steam Coal          ST             SUB              376.0               1962               1964               1960
> 
```

- This tells us that for the plant Utility.ID = 10000 & Plant.Code = 1241
  - Is ~1.5 GW of coal steam turbines (ST)
  - It primarily burns sub bituminous coal
  - Was finished in stages (1973 - 1977)
- For Utility.ID = 10000 & Plant.Code = 2079
  - The NGCC turbine side (CT) is 176 MW, finished in 1997
  - The NGCC steam side (CA) is 146 MW, finished in 2000

Don’t forget Barry!

Our *new* F860 data:

```
> m1[m1$Plant.Code == 3,]
     Utility.ID Plant.Code                       Technology Prime.Mover Energy.Source.1 Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min
2625        195          3          Conventional Steam Coal          ST             BIT             1192.5               1970               1971               1969
2626        195          3 Natural Gas Fired Combined Cycle          CA              NG              390.4               2000               2000               2000
2627        195          3 Natural Gas Fired Combined Cycle          CT              NG              680.4               2000               2000               2000
2628        195          3          Natural Gas Steam Turbine          ST              NG              306.2               1954               1954               1954
> |
```

And our current F923 data:

```
> f923[f923$Plant.Id == 3,][,c(1, 4:6, 14:15, 24:25)]
  Plant.Id Plant.Name     Operator.Name Operator.Id Reported.Prime.Mover Reported.Fuel.Type.Code Net.Generation..Megawatthours. YEAR
1        3      Barry Alabama Power Co         195                   CA                      NG                     2918277.00 2016
2        3      Barry Alabama Power Co         195                   CT                      NG                     5483083.00 2016
3        3      Barry Alabama Power Co         195                   ST                     BIT                     4278313.10 2016
4        3      Barry Alabama Power Co         195                   ST                      NG                       91217.94 2016
>
```

We are now closer in dimensions for the F860 and F923 data, but we are not exact = that is life

```
> dim(m1)
[1] 9392    9
> dim(f923)
[1] 12558    25
> head(m1)
  Utility.ID Plant.Code                        Technology Prime.Mover Energy.Source.1 Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min
1      10000       1241              Conventional Steam Coal          ST             SUB             1578.0               1975               1977               1973
2      10000       2079              Conventional Steam Coal          ST             SUB              594.3               1969               1969               1969
3      10000       2079      Natural Gas Fired Combined Cycle          CA              NG              146.0               2000               2000               2000
4      10000       2079      Natural Gas Fired Combined Cycle          CT              NG              176.0               1997               1997               1997
5      10000       2079 Natural Gas Fired Combustion Turbine          GT              NG              154.0               2000               2000               2000
6      10000       2080              Conventional Steam Coal          ST             SUB              376.0               1962               1964               1960
> f923[1:5, c(1, 4:6, 14:15, 24:25)]
  Plant.Id          Plant.Name     Operator.Name Operator.Id Reported.Prime.Mover Reported.Fuel.Type.Code Net.Generation..Megawatthours. YEAR
1        3               Barry Alabama Power Co         195                   CA                      NG                     2918277.00 2016
2        3               Barry Alabama Power Co         195                   CT                      NG                     5483083.00 2016
3        3               Barry Alabama Power Co         195                   ST                     BIT                     4278313.10 2016
4        3               Barry Alabama Power Co         195                   ST                      NG                       91217.94 2016
5        4 Walter Bouldin Dam Alabama Power Co         195                   HY                     WAT                      468960.00 2016
> |
```

Let's try a merge and see what happens!

```
> m2 <- merge(x = m1, y = f923, by.x = c('Utility.ID', 'Plant.Code', 'Prime.Mover', 'Energy.Source.1'), by.y = c('Operator.Id', 'Plant.Id', 'Reported.Prime.Mover', 'Reported.Fuel.Type.Code'))
> dim(m2)
[1] 8795   30
> m2[1:10,c(1:9, 29:30)]
   Utility.ID Plant.Code Prime.Mover Energy.Source.1                           Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours. YEAR
1       10000       1241          ST             SUB              Conventional Steam Coal             1578.0               1975               1977               1973                      7933023.1 2016
2       10000       2079          CA              NG     Natural Gas Fired Combined Cycle              146.0               2000               2000               2000                        12726.0 2016
3       10000       2079          CT              NG     Natural Gas Fired Combined Cycle              176.0               1997               1997               1997                        45462.0 2016
4       10000       2079          GT              NG Natural Gas Fired Combustion Turbine              154.0               2000               2000               2000                        41833.0 2016
5       10000       2079          ST             SUB              Conventional Steam Coal              594.3               1969               1969               1969                      2583399.5 2016
6       10000       2080          ST             SUB              Conventional Steam Coal              376.0               1962               1964               1960                       721378.7 2016
7       10000       2081          GT             DFO                    Petroleum Liquids              484.0               1975               1977               1972                            0.0 2016
8       10000       2081          IC             DFO                    Petroleum Liquids                2.0               1985               1985               1985                         1317.0 2016
9       10000      56351          WT             WND                 Onshore Wind Turbine              148.5               2008               2010               2006                       408557.0 2016
10      10000      59772          BA             MWH                            Batteries                1.0               2012               2012               2012                            0.0 2016
> sum(m2$Nameplate.Capacity)
[1] 1242016
> sum(m1$Nameplate.Capacity)
[1] 1178148
> dim(m1)
[1] 9392    9
> |
```

:/ what is going on here?

How did that get bigger???

Sometimes you have to try a lot of things (this took me > 3 hours to figure out)

```
> m1[duplicated(m1)]
data frame with 0 columns and 9392 rows
> f923[duplicated(f923)]
data frame with 0 columns and 12558 rows
> m2[duplicated(m2)]
data frame with 0 columns and 8795 rows
```

That is not it…

```
> setdiff(f923$Plant.Id, m1$Plant.Code)
  [1]   171   301   305   420   629   637   727   733   898   991  1010  1032  1308  1385  1409  1472  1587  1695  1720  1723  1784  1961  2150  2171  2289  2549  3686  4259  6225  7304  7549  8812
 [33]  8816  8823  8824  8825  8827  8828  8829  8831  8832  8834  8835  8837  8838  8841  8843  8845  8846  8847  8848  8850  8851  8852  8853  8857  8858  8865  8866  8867  8868  8899 10125 10223
 [65] 10600 10604 10672 10820 50001 50277 50534 50730 50801 50818 50872 50945 54423 54579 54686 54794 54814 54924 54935 55049 55098 55551 55552 55578 56006 56017 56124 56318 56534 56535 56846 57508
 [97] 57932 58194 58442 58466 58518 58648 58967 58968 58969 58970 59253 59282 59452 59488 59915 59929 60098 60099 60114 60115 60116 60229 60232 60337 60463 60605 60607 60625 60647 60694 60699 60817
[129] 60818 60844 60866 60887 60892 60893 60894 60906 60959 60984 99999
> f923[f923$Plant.Id == 2,]
 [1] Plant.Id                      Combined.Heat.And.Power.Plant Nuclear.Unit.Id               Plant.Name                    Operator.Name
 [6] Operator.Id                   Plant.State                   Census.Region                 NERC.Region                   Reserved
[11] NAICS.Code                    EIA.Sector.Number             Sector.Name                   Reported.Prime.Mover          Reported.Fuel.Type.Code
[16] AER.Fuel.Type.Code            Reserved.1                    Reserved.2                    Physical.Unit.Label           Total.Fuel.Consumption.Quantity
[21] Electric.Fuel.Consumption.Quantity Total.Fuel.Consumption.MMBtu Elec.Fuel.Consumption.MMBtu Net.Generation..Megawatthours. YEAR
<0 rows> (or 0-length row.names)
> m1[m1$Plant.Code == 2,]
     Utility.ID Plant.Code         Technology Prime.Mover Energy.Source.1 Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min
2619        195          2 Conventional Hydroelectric          HY             WAT               53.9               1963               1963               1963
> m1[m1$Plant.Code == 171,]
[1] Utility.ID         Plant.Code         Technology         Prime.Mover        Energy.Source.1    Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min
<0 rows> (or 0-length row.names)
> f923[f923$Plant.Id == 171,]
    Plant.Id Combined.Heat.And.Power.Plant Nuclear.Unit.Id Plant.Name         Operator.Name Operator.Id Plant.State Census.Region NERC.Region Reserved NAICS.Code EIA.Sector.Number       Sector.Name
169      171                             N              NA  Mabelvale Entergy Arkansas Inc         814          AR           WSC        SERC       NA         22                 1 Electric Utility
170      171                             N              NA  Mabelvale Entergy Arkansas Inc         814          AR           WSC        SERC       NA         22                 1 Electric Utility
    Reported.Prime.Mover Reported.Fuel.Type.Code AER.Fuel.Type.Code Reserved.1 Reserved.2 Physical.Unit.Label Total.Fuel.Consumption.Quantity Electric.Fuel.Consumption.Quantity
169                   GT                     DFO                DFO         NA         NA             barrels                               0                                  0
170                   GT                      NG                 NG         NA         NA                 mcf                               0                                  0
    Total.Fuel.Consumption.MMBtu Elec.Fuel.Consumption.MMBtu Net.Generation..Megawatthours. YEAR
169                            0                           0                              0 2016
170                            0                           0                              0 2016
```

A bunch of other stuff that wasn't it

Any guesses as to what is going on here?

There were 411 rows that were similar for the F923 data (same 'Plant.Id', 'Operator.Id', 'Reported.Fuel.Type.Code', and 'Reported.Prime.Mover')

```
> f923_dup <- f923[duplicated(f923[c('Plant.Id', 'Operator.Id', 'Reported.Fuel.Type.Code', 'Reported.Prime.Mover')]),]
> dim(f923_dup)
[1] 411  25
> summary(f923_dup)
    Plant.Id      Combined.Heat.And.Power.Plant Nuclear.Unit.Id                          Plant.Name                           Operator.Name  Operator.Id       Plant.State    Census.Region
 Min.   :   46   N:252                          Min.   :2.000   State-Fuel Level Increment:372   State-Fuel Level Increment:372   Min.   : 195   PA      : 34   SAT     :82
 1st Qu.:99999   Y:159                          1st Qu.:2.000   Browns Ferry              :  2   Exelon Nuclear            :  9   1st Qu.:99999   FL      : 32   MAT     :74
 Median :99999                                  Median :2.000   Oconee                    :  2   Duke Energy Carolinas, LLC:  4   Median :99999   CA      : 26   ENC     :65
 Mean   :91069                                  Mean   :2.211   Palo Verde                :  2   Tennessee Valley Authority:  4   Mean   :92911   TX      : 24   WSC     :46
 3rd Qu.:99999                                  3rd Qu.:2.000   Arkansas Nuclear One      :  1   Arizona Public Service Co :  2   3rd Qu.:99999   NY      : 23   WNC     :35
 Max.   :99999                                  Max.   :4.000   Beaver Valley             :  1   Florida Power & Light Co  :  2   Max.   :99999   NC      : 21   PACC    :34
                                                NA's   :373     (Other)                   : 31   (Other)                   : 18                   (Other):251   (Other):75
  NERC.Region    Reserved          NAICS.Code     EIA.Sector.Number                        Sector.Name  Reported.Prime.Mover Reported.Fuel.Type.Code AER.Fuel.Type.Code Reserved.1
        :372   Mode:logical   Min.   :   22   Min.   :1.000    Commercial NAICS Cogen    : 33   ST     :202          NG     :126             NG     :126        Mode:logical
 SERC   : 15   NA's:411       1st Qu.:99999   1st Qu.:1.000    Commercial NAICS Non-Cogen: 17   IC     : 66          DFO    : 79             DFO    : 79        NA's:411
 RFC    : 13                  Median :99999   Median :2.000    Electric Utility          :119   GT     : 56          NUC    : 38             NUC    : 38
 WECC   :  4                  Mean   :90513   Mean   :3.238    Industrial NAICS Cogen    : 84   PV     : 31          SUN    : 31             SUN    : 31
 FRCC   :  2                  3rd Qu.:99999   3rd Qu.:5.000    Industrial NAICS Non-Cogen: 10   CT     : 16          BIT    : 19             COL    : 30
 NPCC   :  2                  Max.   :99999   Max.   :7.000    NAICS-22 Cogen            : 35   CA     : 15          WDS    : 19             WWW    : 25
 (Other):  3                                                   NAICS-22 Non-Cogen        :113   (Other): 25          (Other): 99             (Other): 82
 Reserved.2     Physical.Unit.Label Total.Fuel.Consumption.Quantity Electric.Fuel.Consumption.Quantity Total.Fuel.Consumption.MMBtu Elec.Fuel.Consumption.MMBtu
 Mode:logical              : 93     Min.   :       0                Min.   :      0                    Min.   :       0             Min.   :       0
 NA's:411       barrels    : 89     1st Qu.:       0                1st Qu.:      0                    1st Qu.:       0             1st Qu.:       0
                mcf        :145     Median :      18                Median :      6                    Median :   20021             Median :   13930
                Mcf        :  5     Mean   :  349992                Mean   : 180211                    Mean   :  878825             Mean   :  394961
                short tons: 79      3rd Qu.:   99114                3rd Qu.:  48280                    3rd Qu.:  371501             3rd Qu.:  225958
                                    Max.   :13621573                Max.   :7981975                    Max.   :29030359             Max.   :14399020

 Net.Generation..Megawatthours.      YEAR
 Min.   :       0              Min.   :2016
 1st Qu.:       0              1st Qu.:2016
 Median :    3948              Median :2016
 Mean   :  828483              Mean   :2016
 3rd Qu.:   54033              3rd Qu.:2016
 Max.   :11717178              Max.   :2016
```

I was assuming that the F923 data were consistent, that was a bad assumption, always check…

```
> a1 <- aggregate(x = f923$Net.Generation..Megawatthours., by = list(f923$Plant.Id, f923$Operator.Id, f923$Reported.Prime.Mover, f923$Reported.Fuel.Type.Code), FUN = sum)
> names(a1) <- c('Plant.Id', 'Operator.Id', 'Reported.Prime.Mover', 'Reported.Fuel.Type.Code', 'Net.Generation..Megawatthours.')
> head(a1)
  Plant.Id Operator.Id Reported.Prime.Mover Reported.Fuel.Type.Code Net.Generation..Megawatthours.
1    10593         135                   ST                      AB                       80270.03
2    54908         772                   ST                      AB                        5606.00
3    57046         772                   ST                      AB                           0.00
4       79         986                   ST                      AB                           0.00
5     1131        3203                   ST                      AB                           0.00
6    55867        6306                   ST                      AB                      137714.79
> m3 <- merge(x = m1, y = a1, by.x = c('Utility.ID', 'Plant.Code', 'Prime.Mover', 'Energy.Source.1'), by.y = c('Operator.Id', 'Plant.Id', 'Reported.Prime.Mover', 'Reported.Fuel.Type.Code'))
> dim(m3)
[1] 8756   10
> head(m3)
  Utility.ID Plant.Code Prime.Mover Energy.Source.1                          Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.
1      10000       1241          ST             SUB             Conventional Steam Coal             1578.0               1975               1977               1973                      7933023.1
2      10000       2079          CA              NG        Natural Gas Fired Combined Cycle              146.0               2000               2000               2000                        12726.0
3      10000       2079          CT              NG        Natural Gas Fired Combined Cycle              176.0               1997               1997               1997                        45462.0
4      10000       2079          GT              NG Natural Gas Fired Combustion Turbine              154.0               2000               2000               2000                        41833.0
5      10000       2079          ST             SUB             Conventional Steam Coal              594.3               1969               1969               1969                      2583399.5
6      10000       2080          ST             SUB             Conventional Steam Coal              376.0               1962               1964               1960                       721378.7
> sum(m3$Nameplate.Capacity)
[1] 1153254
>
```

That is much better!

For this dataset, we get 97% of capacity (MW) and 95% of the total net generation (MWh)

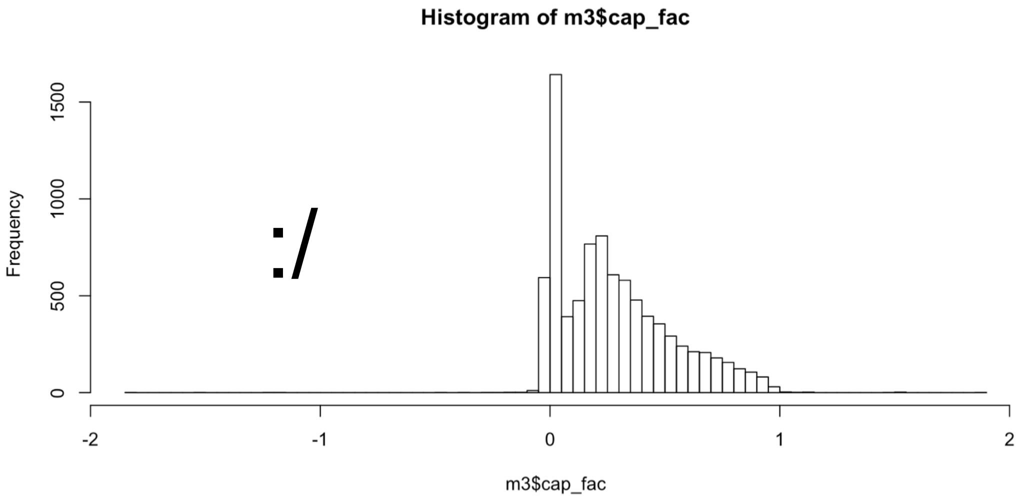
```
> sum(m3$Nameplate.Capacity)/sum(f860$Nameplate.Capacity..MW.)
[1] 0.9788696
> sum(m3$Net.Generation..Megawatthours.)/sum(f923$Net.Generation..Megawatthours.)
[1] 0.9594279
```

- That is not too bad!
- Because the datasets do not perfectly align, it makes some sense that we don't get perfect values
- In this case, we note it to include in future reports, papers, etc., but sometimes good is good enough – but sometimes you have to make that decision

Now, at long last, we can calculate what we set out to do!

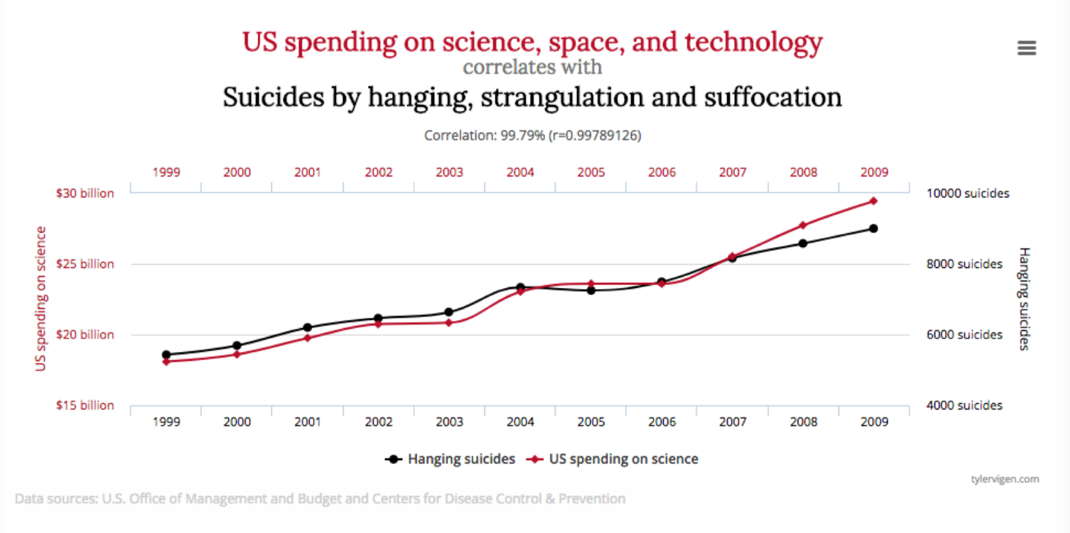
```
> m3$cap_fac <- m3$Net.Generation..Megawatthours./(m3$Nameplate.Capacity*8760)
> head(m3)
  Utility.ID Plant.Code Prime.Mover Energy.Source.1                   Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.     cap_fac
1      10000       1241          ST             SUB      Conventional Steam Coal             1578.0               1975               1977               1973                      7933023.1 0.573888621
2      10000       2079          CA              NG    Natural Gas Fired Combined Cycle       146.0               2000               2000               2000                        12726.0 0.009950272
3      10000       2079          CT              NG    Natural Gas Fired Combined Cycle       176.0               1997               1997               1997                        45462.0 0.029487080
4      10000       2079          GT              NG Natural Gas Fired Combustion Turbine      154.0               2000               2000               2000                        41833.0 0.031009459
5      10000       2079          ST             SUB      Conventional Steam Coal              594.3               1969               1969               1969                      2583399.5 0.496228536
6      10000       2080          ST             SUB      Conventional Steam Coal              376.0               1962               1964               1960                       721378.7 0.219013750
> |
```

```
> summary(m3$cap_fac)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
-1.83357  0.04342  0.22596  0.27538  0.42476  1.85495 
> |
> hist(m3$cap_fac, n = 100)
> |
```



Hmmm, capacity factors < 0 and > 1 should be somewhat suspect

- YOU HAVE TO KNOW YOUR DATA BEFORE YOU MAKE ANY CLAIMS ABOUT IT
- Cue to start diggin’



http://www.tylervigen.com/spurious-correlations

Capacity factors > 1

```
> m3[m3$cap_fac > 1,]
     Utility.ID Plant.Code Prime.Mover Energy.Source.1                  Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.  cap_fac
510       12455      54915          CT              NG Natural Gas Fired Combined Cycle              80.1               1995               1995               1995                      723499.00 1.031101
819       13781       1757          HY             WAT      Conventional Hydroelectric               1.4               1917               1917               1917                       13819.00 1.126794
1744      16653      50639          ST             BLQ         Wood/Wood Waste Biomass              27.6               2001               2001               2001                      277599.31 1.148167
2552      19564      50366          ST              NG        Natural Gas Steam Turbine               4.7               1957               1962               1952                       63388.37 1.539599
2626      19876       3804          CT              NG Natural Gas Fired Combined Cycle             348.0               2003               2003               2003                     3367672.70 1.104706
2645      19876      55939          CT              NG Natural Gas Fired Combined Cycle             892.5               2014               2014               2014                     8526840.00 1.090626
2754      20508      50900          ST             BLQ         Wood/Wood Waste Biomass              34.0               1980               1980               1980                      456679.50 1.533305
3560      34691      50654          BT             GEO                      Geothermal               5.5               2007               2007               2007                       49630.00 1.030095
3575      34691      57281          OT              WH                       All Other               5.0               2010               2010               2010                       46902.00 1.070822
5796      57302      57919          ST             WDS         Wood/Wood Waste Biomass              10.0               1947               1947               1947                      133130.88 1.519759
6452      58627      58696          HY             WAT      Conventional Hydroelectric               1.5               1906               1906               1906                       24374.00 1.854947
7366      59867      10697          ST             BLQ         Wood/Wood Waste Biomass              12.5               1966               1966               1966                      110757.18 1.011481
7613       6028       8907          ST             NUC                         Nuclear            1012.0               1976               1976               1976                     9076168.00 1.023807
8683       9390      54087          ST             BLQ         Wood/Wood Waste Biomass              25.6               1966               1966               1966                      328265.02 1.463796
> sum(m3$Nameplate.Capacity[m3$cap_fac > 1])
[1] 2460.4
> sum(m3$Nameplate.Capacity[m3$cap_fac > 1])/sum(m3$Nameplate.Capacity)
[1] 0.002133442
> sum(m3$Net.Generation..Megawatthours.[m3$cap_fac > 1])
[1] 23198725
> sum(m3$Net.Generation..Megawatthours.[m3$cap_fac > 1])/sum(m3$Net.Generation..Megawatthours.)
[1] 0.005933562
>
```

~0.2% of capacity

~0.6% of energy

- Sometimes nuclear plants run > nameplate
- Some small ones might be bad data

Capacity factors < 0

```
> m3[m3$cap_fac < 0,]
     Utility.ID Plant.Code Prime.Mover Energy.Source.1                      Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.       cap_fac
34        10071      58639          BA             MWH                       Batteries                6.0           2015.000               2015               2015                     -63030.000 -1.199201e+00
64        10210         85          IC             DFO                Petroleum Liquids               25.9           1978.250               1998               1969                      -1111.000 -4.896775e-03
96        10536        506          IC             DFO                Petroleum Liquids               17.2           1953.143               1970               1939                       -948.000 -6.291813e-03
124       10730        507          IC             DFO                Petroleum Liquids                5.6           1950.200               1967               1941                        -89.000 -1.814253e-03
150       11118      57030          IC             DFO                Petroleum Liquids                5.4           2006.000               2006               2006                       -118.000 -2.494504e-03
232       11241       1404          CT              NG  Natural Gas Fired Combined Cycle              125.3           1973.000               1973               1973                      -1347.000 -1.227192e-03

...

1130      14328       6100          PS             WAT     Hydroelectric Pumped Storage             1053.0           1984.000               1984               1984                    -498562.000 -5.404888e-02
1315      15248      60441          BA             MWH                        Batteries                5.0           2013.000               2013               2013                       -472.000 -1.077626e-02
1322      15248       8073          GT              NG Natural Gas Fired Combustion Turbine           24.5           2001.000               2001               2001                        -15.000 -6.989097e-05
1344      15296       2691          PS             WAT     Hydroelectric Pumped Storage             1000.0           1973.000               1973               1973                    -186162.000 -2.125137e-02
1345      15296       2692          PS             WAT     Hydroelectric Pumped Storage              240.0           1961.750               1962               1961                    -284770.000 -1.354500e-01
1440      15452       6156          GT             KER                Petroleum Liquids              181.5           2012.000               2012               2012                       -256.645 -1.614180e-04
1443      15466        467          PS             WAT     Hydroelectric Pumped Storage              300.0           1967.000               1967               1967                    -179267.000 -6.821423e-02
```

- Most of these plants are ~ 0, i.e. they did not produce anything
  - Hotel load draws
- Except batteries and Hydro Pumped Storage!
  - They buy electricity to store, then sell at higher prices
  - Round trip efficiency < 1
  - They are net electricity consumers -> negative capacity factors in our current def’n

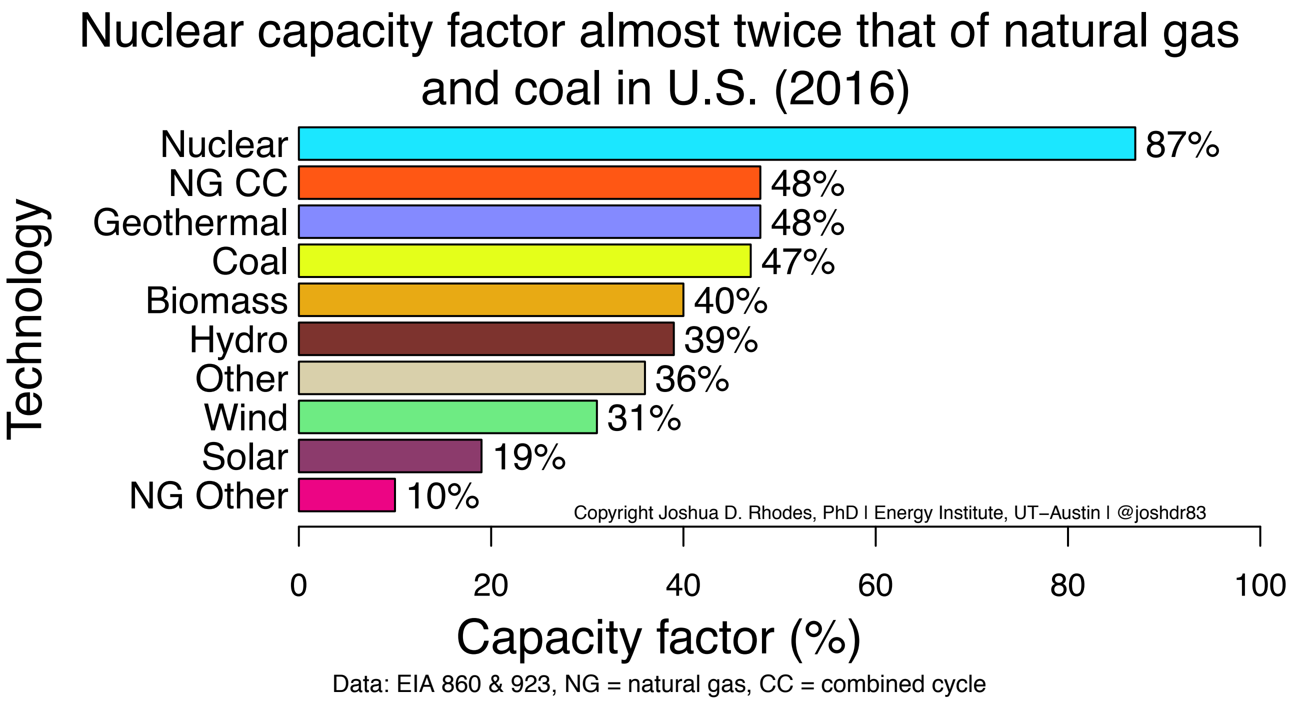
Hydro Pumped Storage (+ Batteries) makes sense for capacity factor to be < 0

```
> m3[m3$Technology == 'Hydroelectric Pumped Storage',]
     Utility.ID Plant.Code Prime.Mover Energy.Source.1                  Technology Nameplate.Capacity Operating.Year.avg Operating.Year.max Operating.Year.min Net.Generation..Megawatthours.      cap_fac
169       11208        392          PS             WAT Hydroelectric Pumped Storage             1626.0           1976.167               1978               1973                         378922  0.026602667
202       11208      57854          PS             WAT Hydroelectric Pumped Storage                5.4           1993.000               1993               1993                              0  0.000000000
931       13994       6189          PS             WAT Hydroelectric Pumped Storage              847.8           1995.000               1995               1995                        -489099 -0.065856593
1130      14328       6100          PS             WAT Hydroelectric Pumped Storage             1053.0           1984.000               1984               1984                        -498562 -0.054048880
1344      15296       2691          PS             WAT Hydroelectric Pumped Storage             1000.0           1973.000               1973               1973                        -186162 -0.021251370
1345      15296       2692          PS             WAT Hydroelectric Pumped Storage              240.0           1961.750               1962               1961                        -284770 -0.135449962
1443      15466        467          PS             WAT Hydroelectric Pumped Storage              300.0           1967.000               1967               1967                        -179267 -0.068214231
1698      16572        145          PS             WAT Hydroelectric Pumped Storage               99.8           1972.000               1972               1972                           3727  0.004263092
1703      16572        148          PS             WAT Hydroelectric Pumped Storage               54.3           1971.000               1971               1971                          12071  0.025376944
1889      17539       6126          PS             WAT Hydroelectric Pumped Storage              587.2           1978.000               1978               1978                        -197011 -0.038300137
1938      17609        104          PS             WAT Hydroelectric Pumped Storage              199.8           1987.000               1987               1987                         238699  0.136380102
2256      18642       2780          PS             WAT Hydroelectric Pumped Storage               95.0           1956.000               1956               1956                              0  0.000000000
2303      18642       6151          PS             WAT Hydroelectric Pumped Storage             1713.6           1978.750               1979               1978                        -703946 -0.046894919
```

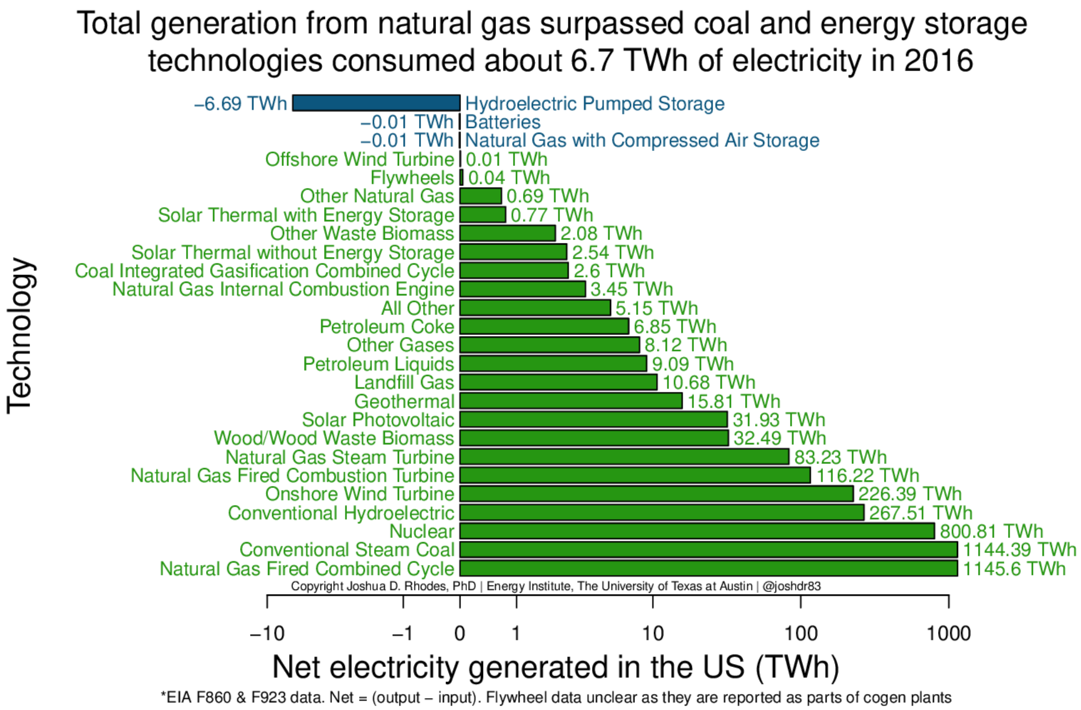
```
> sum(m3$Nameplate.Capacity[m3$cap_fac <= 1 & m3$cap_fac >= 0])/sum(m3$Nameplate.Capacity)
[1] 0.9762014
> sum(m3$Net.Generation..Megawatthours.[m3$cap_fac <= 1 & m3$cap_fac >= 0])/sum(m3$Net.Generation..Megawatthours.)
[1] 0.996165
> |
```

- Not terrible!

With this data, you can do things like this



Or something like this, which I did, for fun



Save all that hard work!

```
> write.csv(x = m3, file = 'merged_2016_f860_f923_data.csv', row.names = F)
> 
```