

# Data for *usensys* models

@olugovoy

June 13, 2019

## Contents

<b>Obtaining EIA data</b>	<b>1</b>
By states . . . . .	1
Power plants & generator level data . . . . .	3
Form EIA-860 detailed data with previous form data (EIA-860A/860B . . . . .	3
Form EIA-923 detailed data with previous form data (EIA-906/920) . . . . .	3
Scraping & downloading all zip and xls files of the forms . . . . .	3
Processing plant-level data . . . . .	6
<b>EPA data</b>	<b>7</b>
EPA's Power Sector Modeling Platform v6 - November 2018 Reference Case . . . . .	7
Get IPM data files . . . . .	7
Coal . . . . .	8
Gas . . . . .	9
<b>tbc</b>	<b>9</b>

## Obtaining EIA data

### By states

Downloading data from <https://www.eia.gov/electricity/data/state/>.

Version:

Final annual data for 2017

Release Date: October 12, 2018

Next Release Date: November 2019

Re-released: January 15, 2019 Revision/Corrections <https://www.eia.gov/electricity/data/state/correction.php>

```
eiadir <- file.path(getwd(), "data/EIA")
if(!dir.exists(eiadir)) dir.create(eiadir, recursive = T)

{www <- list()
# Annual data ####
# 1990 - 2017 Net Generation by State by Type of Producer by Energy Source (EIA-906, EIA-920, and EIA-923)
www$annual_generation_state.xls <- "https://www.eia.gov/electricity/data/state/annual_generation_state.xls"
# 1990 - 2017 Fossil Fuel Consumption for Electricity Generation by Year, Industry Type and State (EIA-906)
www$annual_consumption_state.xls <- "https://www.eia.gov/electricity/data/state/annual_consumption_state.xls"
# 1990 - 2017 Existing Nameplate and Net Summer Capacity by Energy Source, Producer Type and State (EIA-860)
www$existcapacity_annual.xlsx <- "https://www.eia.gov/electricity/data/state/existcapacity_annual.xlsx"
# 2018 - 2022 Proposed Nameplate and Net Summer Capacity by Year, Energy Source, and State (EIA-860)
www$plancapacity_annual.xlsx <- "https://www.eia.gov/electricity/data/state/plancapacity_annual.xlsx"
# 1990 - 2017 U.S. Electric Power Industry Estimated Emissions by State (EIA-767, EIA-906, EIA-920, and EIA-923)
www$emission_annual.xls <- "https://www.eia.gov/electricity/data/state/emission_annual.xls"}
```

```

# 1990 - 2017 Average Price by State by Provider (EIA-861)5
www$avgprice_annual.xlsx <- "https://www.eia.gov/electricity/data/state/avgprice_annual.xlsx"
# 1990 - 2017 Number of Retail Customers by State by Sector (EIA-861)5
www$customers_annual.xlsx <- "https://www.eia.gov/electricity/data/state/customers_annual.xlsx"
# 1990 - 2017 Retail Sales of Electricity by State by Sector by Provider (EIA-861)5
www$sales_annual.xlsx <- "https://www.eia.gov/electricity/data/state/sales_annual.xlsx"
# 1990 - 2017 Revenue from Retail Sales of Electricity by State by Sector by Provider (EIA-861)5
www$revenue_annual.xlsx <- "https://www.eia.gov/electricity/data/state/revenue_annual.xlsx"
# 2001 - 2003 Financial Data on Publicly Owned Electric Utilities with Generation Facilities by State
www$financewgen_annual.xls <- "https://www.eia.gov/electricity/data/state/financewgen_annual.xls"
# 2001 - 2003 Financial Data on Publicly Owned Electric Utilities without Generation Facilities by State
www$financewogen_annual.xls <- "https://www.eia.gov/electricity/data/state/financewogen_annual.xls"

# Monthly data from Electric Power Monthly ####
# 2001 - Present Net Generation by State by Type of Producer by Energy Source1
www$generation_monthly.xlsx <- "https://www.eia.gov/electricity/data/state/generation_monthly.xlsx"
# 2001 - Present Fossil Fuel Consumption for Electricity Generation by Year, Industry Type and State
www$consumption_monthly.xlsx <- "https://www.eia.gov/electricity/data/state/consumption_monthly.xlsx"

download <- function(url, path = eiadir, overwrite = FALSE) {
  fl <- file.path(path, basename(url))
  if(file.exists(fl) & !overwrite) {
    # message("File '", basename(url), "' already exists.")
    return("exists")
  } else {
    download.file(url = url, destfile = fl, mode = "wb")
    if(file.exists(fl)) {
      return("downloaded")
    } else {
      return("download failed")
    }
  }
}

sapply(www, download, overwrite = F)

# Annual capacity
elc_cap <- read_excel(file.path(eiadir, basename(www$existcapacity_annual.xlsx)),
                      sheet = "Existing Capacity",
                      range = "A1:H42756")

elc_cap

# Fuels consumption for electricity generation by months
fue4elc <- read_excel(file.path(eiadir, basename(www$consumption_monthly.xlsx)),
                      sheet = "2017_Preliminary",
                      range = "A5:F9391")

fue4elc

# Electricity generation by months and energy type
elc_gen <- read_excel(file.path(eiadir, basename(www$generation_monthly.xlsx)),
                      sheet = "2017_Preliminary",
                      range = "A5:F24880")

elc_gen

```

```

elc_emis <- read_excel(file.path(eiadir, basename(www$emission_annual.xls)),
                      sheet = "State Emissions",
                      range = "A1:G24880")

elc_emis

save(elc_cap, fue4elc, elc_gen, elc_emis, file = file.path(eiadir, "eia_raw.RData"))

```

## Power plants & generator level data

### Form EIA-860 detailed data with previous form data (EIA-860A/860B)

<https://www.eia.gov/electricity/data/eia860/>

Release Date: September 13, 2018, Final 2017 data

Next release date: June 2019, Early release 2018 data

The survey Form EIA-860 collects generator-level specific information about existing and planned generators and associated environmental equipment at electric power plants with 1 megawatt or greater of combined nameplate capacity.

### Form EIA-923 detailed data with previous form data (EIA-906/920)

<https://www.eia.gov/electricity/data/eia923/>

Monthly (M) release date: April 29, 2019 for February 2019 data

Next monthly release: End of May 2019 (March 2019 data)

Final 2017 data: September 20, 2018

Re-released: March 6, 2019 Corrections/Revisions

Update notice: The moisture content and chlorine content for coal receipts were added to the EIA-923 preliminary data file.

The survey Form EIA-923 collects detailed electric power data – monthly and annually – on electricity generation, fuel consumption, fossil fuel stocks, and receipts at the power plant and prime mover level. Specific survey information provided:

- \* Schedule 2 - fuel receipts and costs

- \* Schedules 3A & 5A - generator data including generation, fuel consumption and stocks

- \* Schedule 4 - fossil fuel stocks

- \* Schedules 6 & 7 - non-utility source and disposition of electricity

- \* Schedules 8A-F - environmental data

- \* Monthly data (M) - over 1,800 plants from the monthly survey

- \* Annual final data - approximately 2,350 monthly plants + 5,790 plants from the annual survey

Summary data

<https://www.eia.gov/electricity/data.php#gencapacity>

### Scraping & downloading all zip and xls files of the forms

(Another option is EIA's API <https://www.eia.gov/opendata/register.php>, though it is not clear yet if the API has all the the data in xls, need to check.)

```

scrap <- function(weblink, filepatern = "([a-zA-Z0-9_]+.zip)|([a-zA-Z0-9_]+.xls)$", aslist = T) {
  # library(rvest)
  links <- read_html(weblink) %>%
    html_nodes("a") %>%
    html_attr('href')
}

```

```

links <- links[grepl(filepatern, links)]
links <- paste0(weblink, links)
if (aslist) {
  lnames <- str_extract(links, filepatern)
  stopifnot(length(links) == length(lnames))
  links <- as.list(links)
  names(links) <- lnames
}
links
}

www <- scrap("https://www.eia.gov/electricity/data/eia860/")
www <- c(www, scrap("https://www.eia.gov/electricity/data/eia923/"))
sapply(www, download, overwrite = F)

fls <- names(www)

# Select data for 2017
fls <- fls[grepl("2017", fls)]
zz <- grepl("zip$", fls) # select zip-files for unzipping
# Unzip files
for (f in fls[zz]) {
  dirname <- gsub("\\.zip$", "", f)
  dirname <- file.path(eiadir, dirname)
  if (!dir.exists(dirname)) dir.create(dirname)
  unzip(file.path(eiadir, f), exdir = dirname, overwrite = T)
}

# dirname <- gsub("\\.zip$", "", fls[1])
dbl2int <- function(x) {
  # coersing unnecesary doubles into integers
  stopifnot(any(class(x) == "data.frame"))
  ii <- sapply(x, is.double)
  # browser()
  for (i in (1:dim(x)[2])[ii]) {
    if (sum(x[[i]] - floor(x[[i]]), na.rm = T) == 0) x[[i]] <- as.integer(x[[i]])
  }
  x
}

```

Reading xls(x) files

```

cleanread_xls <- function(path, sheet, range, dbl2int = T) {
  # Clean read withoug logical values
  xls <- read_excel(path, sheet, range)
  col_type <- sapply(xls, class)
  # browser()
  if (any(col_type == "logical")) { # reimport
    message("Rereading the xls-file, forcing out logical columns.")
    col_type[col_type == "character"] <- "text"
    col_type[col_type == "logical"] <- "text"
    xls <- read_excel(path, sheet, range, col_types = col_type)
  }
  if (dbl2int) xls <- dbl2int(xls)
}

```

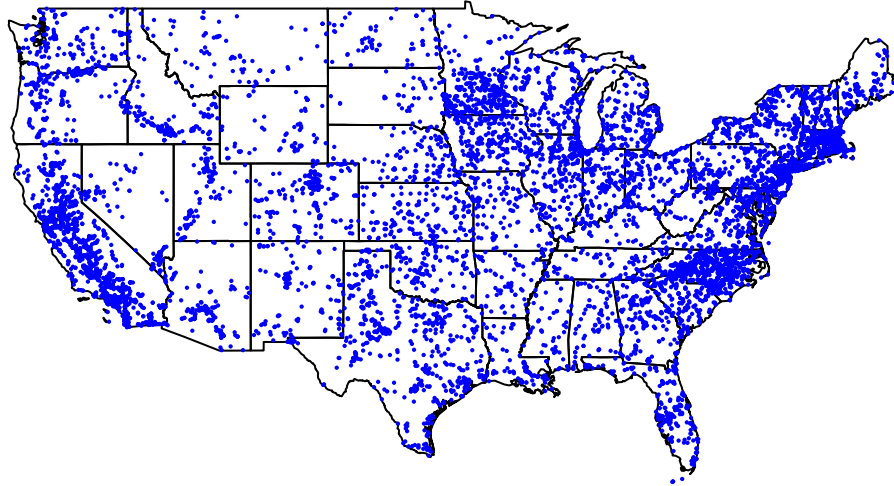


Figure 1: Mapping all power plants.

```

    xls
  }

  elc_pp <- cleanread_xls(file.path(eiadir, "eia8602017/2___Plant_Y2017.xlsx"),
                        sheet = "Plant",
                        range = "A2:AP10129")
  col_type <- sapply(elc_pp, class)
  unique(col_type)

  load("data/maps/usa49reg.RData")
  plot(usa49reg)
  points(elc_pp$Longitude, elc_pp$Latitude, pch = 16, col = "blue", cex = .3)

  save(elc_pp, file = file.path(eiadir, "elc_pp.RData"))

  # continue..
  elc_gn <- cleanread_xls(file.path(eiadir, "eia8602017/3_1_Generator_Y2017.xlsx"),
                        sheet = "Operable",
                        range = "A2:BU21439")
  col_type <- sapply(elc_gn, class)
  unique(col_type)
  elc_gn

```

```

elc_genfue <- cleanread_xls(file.path(eiadir, "f923_2017/EIA923_Schedules_2_3_4_5_M_12_2017_Final_Revis
                                sheet = "Page 1 Generation and Fuel Data",
                                range = "A6:CS13385")

elc_genfue
names(elc_genfue)
col_type <- sapply(elc_genfue, class)
unique(col_type)

# Split annual and monthly data
month_cols <- grepl(paste(month.name, collapse = "|"), names(elc_genfue))
summary(month_cols)

elc_genfue_y <- elc_genfue[, !month_cols]
dim(elc_genfue_y)

jj <- grepl("(Plant.id)|(YEAR)", names(elc_genfue), ignore.case = T)
elc_genfue_m <- bind_cols(elc_genfue[, which(jj)], elc_genfue[, month_cols])
dim(elc_genfue_m)

# reshape monthly data
aa <- gather(elc_genfue_m, key = "key", value = "value", ~`Plant Id`, ~YEAR)
aa$variable <- str_extract(aa$key, "~[a-zA-Z0-9_]+")
unique(aa$variable)
aa$month <- str_extract(aa$key, "[a-zA-Z0-9_]+$")
unique(aa$month)

elc_genfue_m <- select(aa, `Plant Id`, YEAR, month, variable, value)

elc_genfue <- left_join(elc_genfue_m, select(elc_genfue_y, ~starts_with("Reserved")))

clean_names <- function(x) {
  nx <- names(x) %>%
    str_replace_all("\r\n", "_") %>%
    str_replace_all("\r", "_") %>%
    str_replace_all("\n", "_") %>%
    str_replace_all(" ", "_") %>%
    str_replace_all("__", "_") %>%
    str_replace_all("\\$", "USD") %>%
    str_replace_all("/", "_") %>%
    str_replace_all("\\(", "") %>%
    str_replace_all("\\)", "")
  names(x) <- nx
  x
}

elc_genfue <- clean_names(elc_genfue)
names(elc_genfue)

```

## Processing plant-level data

```

# Fuels
elc_genfue_y <- clean_names(elc_genfue_y)

```

```

names(elc_genfue_y)

unique(elc_genfue_y$Reported_Fuel_Type_Code)
unique(elc_genfue_y$AER_Fuel_Type_Code)

unique(elc_gn$Technology)
elc_gn <- clean_names(elc_gn)
names(elc_gn)

aa <- elc_gn %>% group_by(Technology, State) %>%
  summarise(cap = sum(`Nameplate_Capacity_(MW)`)

summary(aa)
sum(aa$cap)/1e3

(unique(aa$Technology))

```

## EPA data

### EPA's Power Sector Modeling Platform v6 - November 2018 Reference Case

“March 21, 2019 - EPA is making the latest power sector modeling platform available, including the associated input data and modeling assumptions, outputs, and documentation.” <https://www.epa.gov/airmarkets/documentation-epas-power-sector-modeling-platform-v6-november-2018-reference-case> Accessed: June 13, 2019

#### Get IPM data files

```

epadir <- file.path(getwd(), "data/EPA")
ipmdir <- file.path(epadir, "IPM")
if(!dir.exists(ipmdir)) dir.create(ipmdir, recursive = T)

ipmweb <- "https://www.epa.gov/airmarkets/documentation-epas-power-sector-modeling-platform-v6-november-2018-reference-case"

scrap_ipm <- function(weblink, filepatern = "([a-zA-Z0-9_]+.zip)|([a-zA-Z0-9_]+.xls(|x))$", aslist = T)
# library(rvest)
links <- read_html(weblink) %>%
  html_nodes("body") %>% html_nodes("section") %>%
  html_nodes("div") %>% html_nodes("div") %>%
  html_nodes("div") %>% html_nodes("ul") %>%
  html_nodes("li") %>% html_nodes("ul") %>%
  html_nodes("li") %>% html_nodes('span') %>%
  html_nodes('a') %>% html_attr('href')
links <- links[grepl(filepatern, links)]
links <- paste0("https://www.epa.gov", links)
if (aslist) {
  lnames <- str_extract(links, filepatern)
  stopifnot(length(links) == length(lnames))
  links <- as.list(links)
  names(links) <- lnames
}

```

```

}
  links
}
www <- scrap_ipm(ipmweb)
sapply(www, download, overwrite = F, path = ipmdir)

```

...

## Coal

```

# Documentation
download(url = "https://www.epa.gov/sites/production/files/2019-03/documents/chapter_7.pdf",
         path = ipmdir, overwrite = F)

# Read coal supply data
coa_sup <- cleanread_xls(file.path(ipmdir, "table_7-26_coal_supply_curves_in_epa_platform_v6.xlsx"),
                        sheet = "Table 7-26 Coal",
                        range = "B4:I9002")
col_type <- sapply(coa_sup, class)
unique(col_type)

coa_sup <- clean_names(coa_sup)
names(coa_sup)
coa_sup

unique(coa_sup$Coal_Grade)
unique(coa_sup$Coal_Supply_Region)
unique(coa_sup$Year)

# Coal type
coa_sup$Coal_Type <- substr(coa_sup$Coal_Grade, 1, 1)
unique(coa_sup$Coal_Type)

# Aggregate by coal types
coa_sup_agg <- coa_sup %>%
  group_by(Year, Coal_Type, Coal_Supply_Region) %>%
  summarise(
    Heat_Content_MMBtu_Ton = mean(Heat_Content_MMBtu_Ton, na.rm = T),
    Cost_of_Production_2016USD_Ton =
      weighted.mean(Cost_of_Production_2016USD_Ton,
                    Coal_Production_Million_Tons_Year, na.rm = T),
    Coal_Production_Million_Tons_Year = sum(Coal_Production_Million_Tons_Year, na.rm = T),
    Coal_Reserves_Million_Tons = sum(Coal_Reserves_Million_Tons, na.rm = T)
  )

coa_type_names <- c(B = "Bituminous", S = "Subbituminous", L = "Lignite")
coa_sup_agg$Coal_Type_Long <- plyr::revalue(coa_sup_agg$Coal_Type, coa_type_names)
# write_csv(coa_sup_agg, path = "tmp/coa_sup_agg.csv")
write_tsv(coa_sup_agg, path = "tmp/coa_sup_agg.tab")

```



## Gas

```
download(url = "https://www.epa.gov/sites/production/files/2019-03/documents/chapter_8.pdf",
         path = ipmdir, overwrite = F)

# Read gas supply curves data
gas_sup_curves <- cleanread_xls(
  file.path(ipmdir, "table_8-5_natural_gas_supply_curves_for_epa_platform_v6.xlsx"),
  sheet = 1,
  range = "B4:D484")

gas_sup_curves <- clean_names(gas_sup_curves)

# Aggregated supply with average prices
gas_sup_avr <- gas_sup_curves %>%
  group_by(Year) %>%
  summarise(Price_2016USD = weighted.mean(Price_2016USD, Natural_Gas_Supply_to_Power_Sector_Quads),
            Natural_Gas_Supply_to_Power_Sector_Quads = sum(Natural_Gas_Supply_to_Power_Sector_Quads))

# Read basis data
gas_sup_basis <- cleanread_xls(
  file.path(ipmdir, "table_8-4_natural_gas_basis_for_epa_platform_v6.xlsx"),
  sheet = 1,
  range = "B4:F596")
gas_sup_basis <- clean_names(gas_sup_basis)

all(unique(gas_sup_basis$Year) == unique(gas_sup_avr$Year)) # Check

# Merge with prices
gas_sup_byreg <- full_join(gas_sup_basis, gas_sup_avr)

write_tsv(gas_sup_avr, path = "tmp/gas_sup_avr.tab")
write_tsv(gas_sup_byreg, path = "tmp/gas_sup_byreg.tab")
```

tbc

...