Data for *usensys* models

@olugovoy

June 07, 2020

Contents

Info

The purpose of the code below is to acquire and preprocess US electric power sector data from EIA, EPA, and other sources. The final processed data, which is required to run USENSYS model, is stored in the '/data' directory in '.RData' format in the model working directory. It is not required to run scripts in this file to be able to run all the versions of USENSYS, available on GitHub. The code below is provided for the reference.

Changelog

April, 2020
- pre-processing data for the USENSYS-IPM
- updated download/read functions

...

June 13, 2019
- initical commit, functions to webscrapp, download, and preprocess the data.

Obtaining EIA data

By states

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

Version:
Final annual data for 2018 Release Date: October 22, 2019 Next Release Date: November 2020

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

{www <- list()
# Annual data ####
# 1990 - 2018   Net Generation by State by Type of Producer by Energy Source (EIA-906, EIA-920, and EIA
www$annual_generation_state.xls <- "https://www.eia.gov/electricity/data/state/annual_generation_state.
# 1990 - 2018   Fossil Fuel Consumption for Electricity Generation by Year, Industry Type and State (EI
www$annual_consumption_state.xls <- "https://www.eia.gov/electricity/data/state/annual_consumption_stat
# 1990 - 2018   Existing Nameplate and Net Summer Capacity by Energy Source, Producer Type and State (E
www$existcapacity_annual.xlsx <- "https://www.eia.gov/electricity/data/state/existcapacity_annual.xlsx"
# 2018 - 2023   Proposed Nameplate and Net Summer Capacity by Year, Energy Source, and State (EIA-860)1
www$plancapacity_annual.xlsx <- "https://www.eia.gov/electricity/data/state/plancapacity_annual.xlsx"
# 1990 - 2018   U.S. Electric Power Industry Estimated Emissions by State (EIA-767, EIA-906, EIA-920, a
www$emission_annual.xls <- "https://www.eia.gov/electricity/data/state/emission_annual.xls"
# 1990 - 2018   Average Price by State by Provider (EIA-861)5
www$avgprice_annual.xlsx <- "https://www.eia.gov/electricity/data/state/avgprice_annual.xlsx"
# 1990 - 2018   Number of Retail Customers by State by Sector (EIA-861)5
www$customers_annual.xlsx <- "https://www.eia.gov/electricity/data/state/customers_annual.xlsx"
# 1990 - 2018   Retail Sales of Electricity by State by Sector by Provider (EIA-861)5
www$sales_annual.xlsx <- "https://www.eia.gov/electricity/data/state/sales_annual.xlsx"
# 1990 - 2018   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 St
```

```
www$financewogen_annual.xls <- "https://www.eia.gov/electricity/data/state/financewogen_annual.xls"

# Monthly data from Electric Power Monthly ####
# Electric Power Monthly with data for November 2019
#   Release date: January 30, 2020
#   Next release date: End of February 2020

# 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"}

sapply(www, download, overwrite = F)

# Annual capacity
elc_cap <- read_excel(file.path(eiadir, basename(www$existcapacity_annual.xlsx)),
                      sheet = "Existing Capacity",
                      range = "A2:H44621")
elc_cap

# Fuels consumption for electricity generation by months
fue4elc <- read_excel(file.path(eiadir, basename(www$consumption_monthly.xlsx)),
                      sheet = "2018_Preliminary",
                      range = "A5:F9361")
fue4elc

# Electricity generation by months and energy type
elc_gen <- read_excel(file.path(eiadir, basename(www$generation_monthly.xlsx)),
                      sheet = "2018_Preliminary",
                      range = "A5:F25156")
elc_gen

elc_emis <- read_excel(file.path(eiadir, basename(www$emission_annual.xls)),
                      sheet = "State Emissions",
                      range = "A1:G41765")
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 3, 2019, Final 2018 data Next release date: June 2020, Early release 2019 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: January 30, 2020 for November 2019 data Next monthly release: End of February 2020 (December 2019 data) Annual release date: September 20, 2019, Final 2018 data 2018 Re-released: January 8, 2020 Corrections/Revisions

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) -approximately 2,026 plants from the monthly survey
Annual final data - approximately 2,026 monthly plants + 7,278 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 2018
fls <- fls[grepl("2018", 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])
```

Reading xls(x) files
Note: reasing from Excel files should be supervised and adjusted in case of any updates of the files on EIA website. Also may not work as expected on some OS/platforms.

```
elc_pp <- cleanread_xls(file.path(eiadir, "eia8602018/2___Plant_Y2018.xlsx"),
                        sheet = "Plant",
                        range = "A2:AP10982")
col_type <- sapply(elc_pp, class)
unique(col_type)
names(elc_pp)
unique(elc_pp$`Liquefied Natural Gas Storage`)
```

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

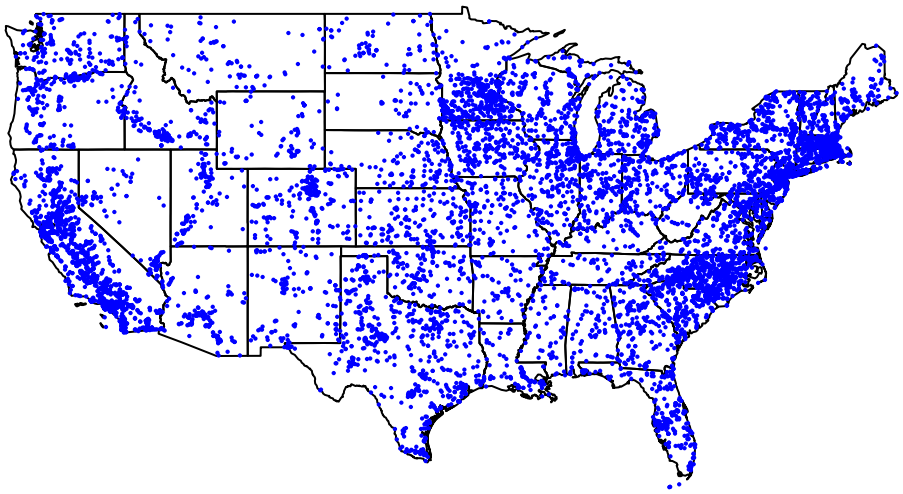

Figure 1: Mapping all power plants.

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

```
# continue..
elc_gn <- cleanread_xls(file.path(eiadir, "eia8602018/3_1_Generator_Y2018.xlsx"),
                        sheet = "Operable",
                        range = "A2:BU22120")
```

```r
col_type <- sapply(elc_gn, class)
unique(col_type)
elc_gn
names(elc_gn)

elc_genfue <- cleanread_xls(file.path(eiadir, "f923_2018/EIA923_Schedules_2_3_4_5_M_12_2018_Final_Revis
                      sheet = "Page 1 Generation and Fuel Data",
                      range = "A6:CS13968")
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")))

elc_genfue <- clean_names(elc_genfue)
names(elc_genfue)
```

Processing plant-level data

```r
# 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")
# needsdir <- 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-

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)

# IPM shape files
download(url = "https://www.epa.gov/sites/production/files/2019-08/ipm_v6_regions.zip",
         overwrite = F, path = ipmdir)
```

```
unzip(file.path(ipmdir, "ipm_v6_regions.zip"),
      exdir = file.path(ipmdir, "ipm_v6_regions"), overwrite = T)
```

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)
save(coa_sup_agg, coa_sup, file = file.path(ipmdir, "coal_supply.RData"))
# 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)
```

```r
# 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)
unique(gas_sup_byreg$IPM_Region)

save(gas_sup_avr, gas_sup_byreg, gas_sup_curves, gas_sup_basis,
     file = file.path(ipmdir, "gas_supply.RData"))

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

NEEDS v6 Database

National Electric Energy Data System (NEEDS) v6 https://www.epa.gov/airmarkets/national-electric-energy-data-system-needs-v6 Accessed: Apr 28, 2020

The National Electric Energy Data System or "NEEDS" database contains the generation unit records used to construct the "model" plants that represent existing and planned/committed units in EPA modeling applications of IPM. NEEDS includes basic geographic, operating, air emissions, and other data on these generating units. NEEDS v6 is a complete bottom-up update for EPA's new power sector modeling platform v6. For a description of the sources used in preparing NEEDS v6, see Documentation, Chapter 4: Generating Resources.

```r
needsweb <- "https://www.epa.gov/airmarkets/national-electric-energy-data-system-needs-v6"

# Most Recent Version of NEEDS:
# NEEDS v6 rev: 3-26-2020(5 MB)
# Comment and Change Log rev: 3-26-2020(142 K)
download(url = "https://www.epa.gov/sites/production/files/2020-04/needs_v6_03-26-2020.xlsx",
         path = ipmdir, overwrite = F)
```

```
# NEEDS v6 Database(4 MB)  (rev: 5-31-2019)
# download(url = "https://www.epa.gov/sites/production/files/2019-06/needs_v6_may_2019_reference_case.x
#          path = ipmdir, overwrite = F)


# Comment and Change Log(88 K) (rev: 5-31-2019)

# NEEDS v6 Database(3 MB)  (rev: 11-30-2018)
# download(url = "https://www.epa.gov/sites/production/files/2019-03/needs_v6_november_2018_reference_c
#          path = ipmdir, overwrite = F)

# Incremental Documentation for NOx Rates in NEEDS v6(3 pp, 44 K) (9-14-2018)

# NEEDS v6 Database(3 MB) (rev: 5-30-2018)


scrap_needs <- 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_needs(needsweb)
sapply(www, download, overwrite = F, path = ipmdir)
```

```
needs <- cleanread_xls(file.path(ipmdir, "needs_v6_03-26-2020.xlsx"),
                       sheet = "NEEDS v6_active",
                       range = "A1:AX18252")
col_type <- sapply(needs, class)
unique(col_type)
needs
names(needs)
save(needs, file = file.path(ipmdir, "NEEDS_v6_active.RData"))
```

Wind potential

Table 4-38 Onshore Regional Potential Wind Capacity (MW) by TRG and Cost Class in EPA Platform v6

```
tbl_4_38_pth <- file.path(
  ipmdir,
```

```
  "table_4-38_onshore_regional_potential_wind_capacity_mw_by_trg_and_cost_class_in_epa_platform_v6.xlsx
file.exists(tbl_4_38_pth)

rename_fill_tbl <- function(tbl) {
  nms <- names(tbl)
  nms <- c(nms[1:3], paste0("Cost_Class_", 1:(ncol(tbl)-3)))
  names(tbl) <- nms
  tbl <- tbl[-1,]

  # fill-in NAs
  r_1 <- NA; s_1 <- NA
  for(i in 1:nrow(tbl)) {
    r <- tbl$IPM_Region[i]
    s <- tbl$State[i]
    if (!is.na(r)) r_1 <- r else tbl$IPM_Region[i] <- r_1
    if (!is.na(s)) s_1 <- s else tbl$State[i] <- s_1
  }
  tbl
}

reshape_MW_tbl <- function(tbl, val_to_name = "MW") {
  tbl <- tbl %>%
  pivot_longer(cols = starts_with("Cost_Class"), names_prefix = "Cost_Class_",
               names_to = "Cost_Class", values_to = val_to_name) %>%
  mutate(Cost_Class = as.integer(Cost_Class))
  tbl
}

onshore_wind_potential <- cleanread_xls(tbl_4_38_pth, "Table 4-38", skip = 3) %>%
  clean_names() %>% rename_fill_tbl() %>% reshape_MW_tbl()
onshore_wind_potential

save(onshore_wind_potential,
     file = file.path(ipmdir, "onshore_wind_potential.RData"))
rm(tbl_4_38_pth, onshore_wind_potential)
```

Wind Generation Profiles

in EPA Platform v6 (kWh of Generation per MW of Capacity)

```
tbl_4_39_pth <- file.path(
  ipmdir, "table_4-39_wind_generation_profiles_in_epa_platform_v6",
  "Table 4-39 Wind Generation Profiles in EPA Platform v6.xlsx")
file.exists(tbl_4_39_pth)
# a <- read_excel(tbl_4_39_pth, "Onshore", skip = 2, col_names = T)

reshape_genpro_tbl <- function(x) {
  pivot_longer(x, cols = starts_with("Hour"), names_prefix = "Hour",
               names_to = "hour", values_to = "kWh_per_MW") %>%
  mutate(hour = as.integer(hour) - 1L)
}

wind_onshore <- cleanread_xls(tbl_4_39_pth, "Onshore", skip = 3) %>%
```

```
  clean_names() %>% reshape_genpro_tbl()

wind_offshore_shallow <- cleanread_xls(
  tbl_4_39_pth, "Offshore Shallow", skip = 3) %>%
  clean_names() %>% reshape_genpro_tbl()

wind_offshore_middepth <- cleanread_xls(
  tbl_4_39_pth, "Offshore Mid-Depth", skip = 3) %>%
  clean_names() %>% reshape_genpro_tbl()

wind_offshore_deep <- cleanread_xls(
  tbl_4_39_pth, "Offshore Deep", skip = 3) %>%
  clean_names() %>% reshape_genpro_tbl()

save(wind_onshore, wind_offshore_shallow,
     wind_offshore_middepth, wind_offshore_deep,
     file = file.path(ipmdir, "wind_profiles.RData"))
rm(tbl_4_39_pth, wind_onshore, wind_offshore_shallow,
   wind_offshore_middepth, wind_offshore_deep)
```

Wind Capital Cost Adder

Table 4-40 Capital Cost Adder (2016$/kW) for New Onshore Wind Plants by Resource and Cost Class in EPA Platform v6

```
tbl_4_40_pth <- file.path(
  ipmdir,
  "table_4-40_capital_cost_adder_for_new_onshore_wind_plants_in_epa_platform_v6.xlsx")
file.exists(tbl_4_40_pth)

onshore_invcost_adder <- cleanread_xls(tbl_4_40_pth, "Table 4-40", skip = 3) %>%
  clean_names() %>% rename_fill_tbl() %>%
  reshape_MW_tbl(val_to_name = "USD_per_kW")
onshore_invcost_adder

save(onshore_invcost_adder,
     file = file.path(ipmdir, "onshore_invcost_adder.RData"))
rm(tbl_4_40_pth, onshore_invcost_adder)
```

Solar potential

Photovoltaic Table 4-41 Solar Photovoltaic Regional Potential Capacity (MW) by Resource and Cost Class in EPA Platform v6

```
tbl_4_41_pth <- file.path(
  ipmdir,
  "table_4-41_solar_photovoltaic_regional_potential_capacity_mw_by_resource_and_cost_class_in_epa_platf

if (file.exists(tbl_4_41_pth)) {
  solar_pv_potential <- cleanread_xls(tbl_4_41_pth, "Table 4-41", skip = 3) %>%
    clean_names() %>% rename_fill_tbl() %>% reshape_MW_tbl()
```

```
  save(solar_pv_potential,
       file = file.path(ipmdir, "solar_pv_potential.RData"))
  rm(tbl_4_41_pth, solar_pv_potential)
}
```

Solar Thermal Table 4-42 Solar Thermal Regional Potential Capacity (MW) by Resource and Cost Class in EPA Platform v6

```
tbl_4_42_pth <- file.path(
  ipmdir,
  "table_4-42_solar_thermal_regional_potential_capacity_mw_by_resource_and_cost_class_in_epa_platform_v6

if (file.exists(tbl_4_42_pth)) {
  solar_th_potential <- cleanread_xls(tbl_4_42_pth, "Table 4-42", skip = 3) %>%
  clean_names() %>% rename_fill_tbl() %>% reshape_MW_tbl()

  save(solar_th_potential,
       file = file.path(ipmdir, "solar_th_potential.RData"))
  rm(tbl_4_42_pth, solar_th_potential)
}
```

Solar generation profiles

Table 4-43 Solar Photovoltaic Generation Profiles in EPA Platform v6 (kWh of Generation per MW of Capacity)

```
tbl_4_43_pth <- file.path(
  ipmdir,
  "table_4-43_solar_photovoltaic_generation_profiles_in_epa_platform_v6.xlsx")

if (file.exists(tbl_4_43_pth)) {
  solar_profiles <- cleanread_xls(tbl_4_43_pth, "Table 4-43", skip = 3) %>%
    clean_names() %>% reshape_genpro_tbl()
  solar_profiles

  save(solar_profiles, file = file.path(ipmdir, "solar_profiles.RData"))
  rm(tbl_4_43_pth, solar_profiles)
}
```

Solar Capital costs adder

Photovoltaic Table 4-44 Solar Photovoltaic Regional Capital Cost Adder (2016$/kW) for Potential Units by Resource and Cost Class in EPA Platform v6

```
tbl_4_44_pth <- file.path(
  ipmdir,
  "table_4-44_capital_cost_adder_for_new_solar_pv_plants_in_epa_platform_v6.xlsx")
file.exists(tbl_4_44_pth)

solpv_invcost_adder <- cleanread_xls(tbl_4_44_pth, "Table 4-44", skip = 3) %>%
  clean_names() %>% rename_fill_tbl() %>%
  reshape_MW_tbl(val_to_name = "USD_per_kW")
```

```
solpv_invcost_adder

save(solpv_invcost_adder,
     file = file.path(ipmdir, "solpv_invcost_adder.RData"))
rm(tbl_4_44_pth, solpv_invcost_adder)
```

Solar Thermal Table 4-45 Solar Thermal Regional Capital Cost Adder (2016$/kW) for Potential Units by Resource and Cost Class in EPA Platform v6

```
tbl_4_45_pth <- file.path(
  ipmdir,
  "table_4-45_capital_cost_adder_for_new_solar_thermal_plants_in_epa_platform_v6.xlsx")
file.exists(tbl_4_45_pth)

solth_invcost_adder <- cleanread_xls(tbl_4_45_pth, "Table 4-45", skip = 3) %>%
  clean_names() %>% rename_fill_tbl() %>%
  reshape_MW_tbl(val_to_name = "USD_per_kW")
solth_invcost_adder

save(solth_invcost_adder,
     file = file.path(ipmdir, "solth_invcost_adder.RData"))
rm(tbl_4_45_pth, solth_invcost_adder)
```

Load duration curves

Table 2-2 2021 Load Duration Curves Used in EPA Platform v6

```
tbl_2_2_pth <- file.path(
  ipmdir,
  "table_2-2_load_duration_curves_used_in_epa_platform_v6.xlsx")
file.exists(tbl_2_2_pth)

load_curves <- cleanread_xls(tbl_2_2_pth, "Table 2-2", skip = 3) %>%
  clean_names() %>%
  pivot_longer(cols = starts_with("Hour"), names_prefix = "Hour_",
               names_to = "hour", values_to = "MWh") %>%
  mutate(hour = as.integer(hour) - 1L)

save(load_curves, file = file.path(ipmdir, "load_curves.RData"))
rm(tbl_2_2_pth, load_curves)
```

Transmission capabilities

Table 3-21 Annual Transmission Capabilities of U.S. Model Regions in EPA Platform v6 - 2021

```
tbl_3_21_pth <- file.path(
  ipmdir,
  "table_3-21_annual_transmission_capabilities_of_u.s._model_regions_in_epa_platform_v6_-_2021.xlsx")
file.exists(tbl_3_21_pth)

transmission <- cleanread_xls(tbl_3_21_pth, "Table 3-21", skip = 3) %>%
  clean_names()
```

```r
# fill-in NAs in "from" column
transmission$From2 <- transmission$From
v_1 <- NA
for(i in 1:nrow(transmission)) {
  v <- transmission$From[i]
  if (!is.na(v)) v_1 <- v
  transmission$From2[i] <- v_1
}
save(transmission, file = file.path(ipmdir, "transmission.RData"))
rm(tbl_3_21_pth, transmission)
```

eGRID

...

NREL

https://atb.nrel.gov/electricity/2018/summary.html
https://atb.nrel.gov/electricity/2019/data.html (included in PowerGenome database)

FERC

https://www.ferc.gov/docs-filing/forms/form-1/data.asp
(included in PUDL database)

PUDL

The Public Utility Data Liberation (PUDL) project is an assembly of EIA, EPA, and FERC forms, standardized and documented.

PowerGenome

https://github.com/gschivley/PowerGenome
A tool to create input files for power system optimization models with an integrated database (a selection from PUDL (EIA, EPA, FERC), NREL, and other sources).
The script in the chunk below reads PowerGenome SQLite-database, re-saves it in R-format, also takes a glimpse at some tables in the database. To run the script, download the *pudl_updated.sqlite* (see the link at https://github.com/gschivley/PowerGenome).