

Script Description Header

File Name: DGStats Installed Capacity by Climate Zone.R

File Location: "~/Desktop/Avoided Cost Calculator to Net Billing Tariff Export Compensation Rate/DGStats Installed Capacity by Climate Zone"

Project: Avoided Cost Calculator to Net Billing Tariff Export Compensation Rate

Description: Calculates Installed Capacity by Climate Zone from DGStats data.

Load Packages

```
library(tidyverse)
library(lubridate)
library(openxlsx)
```

Disable Scientific Notation
options(scipen = 999)

Turn off a confusing message when summarizing data
options(dplyr.summarise.inform = FALSE)

Set Working Directories
setwd("~/Desktop/Avoided Cost Calculator to Net Billing Tariff Export Compensation Rate/DGStats Installed Capacity by Climate Zone")
Code_WD <- getwd()

Load ZIP Code to Climate Zone Mapping

California Energy Commission - Building Climate Zones by ZIP Code Mapping:

<https://www.energy.ca.gov/media/3560>

https://www.energy.ca.gov/sites/default/files/2020-04/BuildingClimateZonesByZIPCode_ada.xlsx

Note: this CEC climate zone mapping includes climate zones 1 through 16.

The Avoided Cost Calculator divides Climate Zone 3 into 3A and 3B.

```
zipCodetoClimateZone <- read.xlsx("https://www.energy.ca.gov/sites/default/files/2020-04/BuildingClimateZonesByZIPCode_ada.xlsx")
```

Note: some ZIP codes are missing from this spreadsheet.

Manually identified missing ZIP codes and found the corresponding climate zones using

[https://caenergy.maps.arcgis.com/apps/webappviewer/index.html?](https://caenergy.maps.arcgis.com/apps/webappviewer/index.html?id=5cfefd9798214bea91cc4fddaa7e643f)

[id=5cfefd9798214bea91cc4fddaa7e643f](https://caenergy.maps.arcgis.com/apps/webappviewer/index.html?id=5cfefd9798214bea91cc4fddaa7e643f)

(CEC EZ Building Climate Zone Finder, last updated 2018-01-03)

```
missingZipCodetoClimateZone <- read.csv("MissingBuildingClimateZonesByZIPCode.csv")
```

```
zipCodetoClimateZone <- rbind(zipCodetoClimateZone, missingZipCodetoClimateZone)
```

```
rm(missingZipCodetoClimateZone)
```

Convert CZ from numeric to character to accommodate creation of "3A" and "3B".

```
zipCodetoClimateZone <- zipCodetoClimateZone %>%
  mutate(Building.CZ = as.character(Building.CZ))
```

Divide Climate Zone 3 into Climate Zones 3A and 3B.

#

<https://www.pge.com/includes/docs/pdfs/about/rates/rebateprogrameval/advisorygroup/climatezones.pdf>

(PDF pgs. 21 - 23)

Converted from PDF to CSV using <https://tabula.technology>

```
zipCodesClimateZone3 <- read.csv("Tabula - PG&E Zones 3A and 3B Distinction.csv") %>%
  select(ZIP, CZ)
```

Note: some ZIP codes are missing from PG&E's 3A/3B spreadsheet

(which was published in 2001).

Manually identified missing ZIP codes and found the corresponding CZs

by looking at adjacent ZIP codes that were in PG&E's dataset.

```
missingZipCodesClimateZone3 <- read.csv("Missing PG&E Zone 3 ZIP Code Mappings.csv")
```

```
zipCodesClimateZone3 <- rbind(zipCodesClimateZone3, missingZipCodesClimateZone3)
```

```
rm(missingZipCodesClimateZone3)
```

Replace CEC's Climate Zone 3 ZIP Code Mapping with PG&E's 3A/3B Mapping.

```

zipCodetoClimateZone <- zipCodetoClimateZone %>%
  filter(Building.CZ != "3")

zipCodesClimateZone3 <- zipCodesClimateZone3 %>%
  rename(Zip.Code = ZIP,
         Building.CZ = CZ)

# There are 7 zip codes that appear in the PG&E 3A/3B mapping,
# and appear in the CEC mapping as not being in Climate Zone 3:
# 94525 (12/3A), 94553 (12/3A), 94565 (12/3A), 94569 (12/3A),
# 94591 (12/3B), 94923 (1/3B), and 95004 (4/3B).
# Zip code boundaries do change over time;
# based on manual review using the CEC EZ Building Climate Zone Finder,
# it appears that the PG&E mapping (published in 2001)
# is more likely to be incorrect,
# and that the CEC mapping (published in 2018)
# is the correct one to use for these 7 zip codes.
zipCodesClimateZone3 <- zipCodesClimateZone3 %>%
  filter(!(Zip.Code %in% zipCodetoClimateZone$Zip.Code))

zipCodetoClimateZone <- rbind(zipCodetoClimateZone, zipCodesClimateZone3)
rm(zipCodesClimateZone3)

#### Download DGStats Data ####
# California Distributed Generation Statistics
# (Download Interconnected Project Sites Data Set):
# https://www.californiadgstats.ca.gov/downloads/
# https://www.californiadgstats.ca.gov/download/interconnection_rule21_projects/

DGStatsDirectory <- tempdir()
DGStatsZip <- tempfile("file", DGStatsDirectory, ".zip")
download.file("https://www.californiadgstats.ca.gov/download/interconnection_rule21_projects/",
             DGStatsZip)
DGStatsFiles <- utils::unzip(DGStatsZip, exdir = DGStatsDirectory, overwrite=T)

#### Load and Clean DGStats Data ####

# Iterate through PG&E, SCE, and SDG&E DGStats files
Utility_Names <- c("PGE", "SCE", "SDGE")

for(Utility_Name_Iter in Utility_Names){

  # Create utility-specific filename and read DGStats file
  # (takes about 15-30 seconds per file, depending on the utility).
  # Filter to just distributed generation projects that are on a NEM tariff.
  # In future years, will need to also include Net Billing Tariff
  # in addition to NEM 1.0 and 2.0.
  dgStatsFilepath <- file.path(DGStatsDirectory,
                              paste0(Utility_Name_Iter,
                                     "_Interconnected_Project_Sites_2022-12-31.csv"))

  dgStatsRaw <- read.csv(dgStatsFilepath) %>%
    filter(NEM.Tariff %in% c("1", "2"))

  rm(dgStatsFilepath)

  # Note: the SCE data include 4115 project sites where System.Size.AC is NA.
  # All of these have a Technology.Type of "Energy Storage",
  # so it's possible to use the values from the "Storage.Size.kW.AC." column.
  # This requires saving these projects to their own variable,
  # filling in the System.Size.AC column,
  # and then adding them back to the main dataframe.
  dgStatsMissingCapacity <- dgStatsRaw %>%
    filter(is.na(System.Size.AC))

  if(nrow(dgStatsMissingCapacity) > 0){
    dgStatsRaw <- dgStatsRaw %>%
      filter(!is.na(System.Size.AC))
  }
}

```

```

dgStatsMissingCapacity <- dgStatsMissingCapacity %>%
  mutate(System.Size.AC = ifelse(Technology.Type == "Energy Storage",
    `Storage.Size..kW.AC.` , System.Size.AC))

dgStatsRaw <- rbind(dgStatsRaw, dgStatsMissingCapacity)
}

rm(dgStatsMissingCapacity)

# Select just the columns that are needed
# to calculate total installed capacity by Climate Zone.
# Some of the county names are in "UPPER CASE" or "lower case".
# Convert all county names to "Title Case" for consistency when summarizing.
dgStatsRaw <- dgStatsRaw %>%
  select(Service.Zip, Service.County, System.Size.AC) %>%
  mutate(Service.County = str_to_title(Service.County))

#### Summarize Installed Capacity by ZIP Code/County ####
# Some project sites have a ZIP Code of "NA".
# There is a mapping of ZIP codes to Climate Zones,
# but some counties have multiple Climate Zones.
# This means that sites with an "NA" ZIP Code
# need to be saved to their own variable,
# and require a different approach to map their capacity to a Climate Zone.

dgStatsZIP <- dgStatsRaw %>%
  filter(!is.na(Service.Zip)) %>%
  group_by(Service.Zip, Service.County) %>%
  summarize(Total_Capacity_kW_AC = sum(System.Size.AC)) %>%
  ungroup()

dgStatsNoZIP <- dgStatsRaw %>%
  filter(is.na(Service.Zip)) %>%
  group_by(Service.County) %>%
  summarize(Total_Capacity_kW_AC = sum(System.Size.AC)) %>%
  ungroup()

rm(dgStatsRaw)

#### Calculate Installed Capacity by CZ for Sites with ZIP Code ####
# It appears that there are some project sites
# in the SCE dataset with incorrect zip codes.
# The County data is in SCE service territory,
# but the ZIP codes are in PG&E or SDG&E service territory,
# and correspond to Climate Zones 2, 3A, 3B, 5, 7, 11, or 12.
# The total amount of capacity associated with these ZIP codes
# is relatively small (874 kW-AC),
# and is set to 0 kW-AC when calculating capacity weighting values
# for the ACC distribution capacity costs.

dgStatsZIP_CZ <- dgStatsZIP %>%
  left_join(zipCodetoClimateZone, by = c("Service.Zip" = "Zip.Code")) %>%
  group_by(Service.County, Building.CZ) %>%
  summarize(Total_Capacity_kW_AC = sum(Total_Capacity_kW_AC)) %>%
  ungroup()

rm(dgStatsZIP)

#### Estimate Installed Capacity by CZ for Sites without ZIP Code ####
# For sites with ZIP code data provided,
# get the percentage of projects of projects in each county
# that fall into each of the county's climate zones.
# Assign capacity from sites without ZIP code information
# (only county information) to climate zones based on those same percentages.

dgStatsCountyWeighting <- dgStatsZIP_CZ %>%

```

```

group_by(Service.County) %>%
mutate(Capacity_Weighting = Total_Capacity_kW_AC /
      sum(Total_Capacity_kW_AC)) %>%
ungroup() %>%
select(Service.County, Building.CZ, Capacity_Weighting)

dgStatsNoZIP_CZ <- dgStatsCountyWeighting %>%
  right_join(dgStatsNoZIP, by = "Service.County") %>% # Only counties in the No-ZIP dataset.
  mutate(Total_Capacity_kW_AC = Total_Capacity_kW_AC * Capacity_Weighting) %>%
  select(Service.County, Building.CZ, Total_Capacity_kW_AC)

rm(dgStatsCountyWeighting, dgStatsNoZIP)

#### Create and Save Final Table of Installed Capacity by Climate Zone in Service Territory ####
# Combine data from sites with ZIP code data with sites without ZIP code data.
# Take sum across climate zones, dropping county-level grouping in the process.

dgStatsCZ <- rbind(dgStatsZIP_CZ, dgStatsNoZIP_CZ) %>%
  group_by(Building.CZ) %>%
  summarize(Total_Capacity_kW_AC = sum(Total_Capacity_kW_AC)) %>%
  ungroup() %>%
  mutate(Building.CZ = factor(Building.CZ,
                              levels = c("1", "2", "3A", "3B",
                                           as.character(seq(4,16)))) %>%

  arrange(Building.CZ)

rm(dgStatsZIP_CZ, dgStatsNoZIP_CZ)

# Create dataframe with all 17 Avoided Cost Calculator climate zones,
# with values of 0 kW-AC for climate zones
# that are not included in the DGStats data for the selected utility.
# Note: some of the project sites in the DGStats database
# appear to have been labeled with the wrong ZIP Code/County information
# and are therefore in Climate Zones
# that are outside of their utility's service territory.
# This is addressed when calculating the Climate Zone weights
# for the Net Billing tariff, but left as-is here.
dgStatsCompleteCZ <- data.frame(Building.CZ = c("1", "2", "3A", "3B",
                                                as.character(seq(4,16)))) %>%

  mutate(Building.CZ = factor(Building.CZ,
                              levels = c("1", "2", "3A", "3B",
                                           as.character(seq(4,16)))) %>%

  left_join(dgStatsCZ, by = "Building.CZ") %>%
  mutate(Total_Capacity_kW_AC = ifelse(is.na(Total_Capacity_kW_AC),
                                      0, Total_Capacity_kW_AC))

rm(dgStatsCZ)

dgStatsCompleteCZ <- dgStatsCompleteCZ %>%
  rename(`E3 ACC Climate Zone` = Building.CZ,
        `Installed NEM & NBT Capacity (kW-AC)` = Total_Capacity_kW_AC)

write.csv(dgStatsCompleteCZ,
          paste0(Utility_Name_Iter,
                " Installed NEM & NBT Capacity by Climate Zone 2022.csv"),
          row.names = F)

}

#### Delete the Temporary Directory Used for DGStats Data ####

unlink(DGStatsZip)
unlink(DGStatsDirectory)

```