

Assignment 1

Rex Shen

9/23/2021

```
# Load Packages
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.2      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(ggplot2)

### Get Electricity Together

Years <- 2017:2020
Quarters <- 1:4
Type <- "Electric"

PGE_Elec <- NULL

for (Year in Years) {
  for (Quarter in Quarters) {

    Filename <- paste0("PGE/Electric/PGE_",
      Year, "_Q", Quarter, "_", Type,
      "UsageByZip.csv")

    Temp <- read_csv(Filename)
    PGE_Elec <- rbind(PGE_Elec, Temp)

  }
}

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
```

```

##  TOTALCUSTOMERS = col_number(),
##  TOTALKWH = col_number(),
##  AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),

```

```

##  AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##  ZIPCODE = col_double(),
##  MONTH = col_double(),
##  YEAR = col_double(),
##  CUSTOMERCLASS = col_character(),
##  COMBINED = col_character(),
##  TOTALCUSTOMERS = col_number(),
##  TOTALKWH = col_number(),
##  AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##  ZIPCODE = col_double(),
##  MONTH = col_double(),
##  YEAR = col_double(),
##  CUSTOMERCLASS = col_character(),
##  COMBINED = col_character(),
##  TOTALCUSTOMERS = col_number(),
##  TOTALKWH = col_number(),
##  AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##  ZIPCODE = col_double(),
##  MONTH = col_double(),
##  YEAR = col_double(),
##  CUSTOMERCLASS = col_character(),
##  COMBINED = col_character(),
##  TOTALCUSTOMERS = col_number(),
##  TOTALKWH = col_number(),
##  AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##  ZIPCODE = col_double(),
##  MONTH = col_double(),
##  YEAR = col_double(),
##  CUSTOMERCLASS = col_character(),
##  COMBINED = col_character(),
##  TOTALCUSTOMERS = col_number(),
##  TOTALKWH = col_number(),
##  AVERAGEKWH = col_number()
## )

```

```

##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##

```

```

## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )

Quarters = 1:2

for (Quarter in Quarters) {

  Filename <- paste0("PGE/Electric/PGE_2021_Q",
    Quarter, "_", Type, "UsageByZip.csv")

  Temp <- read_csv(Filename)
  PGE_Elec <- rbind(PGE_Elec, Temp)
}

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),

```

```

## CUSTOMERCLASS = col_character(),
## COMBINED = col_character(),
## TOTALCUSTOMERS = col_number(),
## TOTALKWH = col_number(),
## AVERAGEKWH = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALKWH = col_number(),
##   AVERAGEKWH = col_number()
## )

PGE_Elec = data.frame(PGE_Elec)

# Get Gas Together

Years <- 2017:2020
Quarters <- 1:4
Type <- "Gas"

PGE_Gas <- NULL

for (Year in Years) {
  for (Quarter in Quarters) {

    Filename <- paste0("PGE/Gas/PGE_",
      Year, "_Q", Quarter, "_", Type,
      "UsageByZip.csv")

    Temp <- read_csv(Filename)
    PGE_Gas <- rbind(PGE_Gas, Temp)
  }
}

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )
##

```

```

##
## -- Column specification -----
## cols(
##     ZIPCODE = col_double(),
##     MONTH = col_double(),
##     YEAR = col_double(),
##     CUSTOMERCLASS = col_character(),
##     COMBINED = col_character(),
##     TOTALCUSTOMERS = col_number(),
##     TOTALTHM = col_number(),
##     AVERAGETHM = col_number()
## )

##
## -- Column specification -----
## cols(
##     ZIPCODE = col_double(),
##     MONTH = col_double(),
##     YEAR = col_double(),
##     CUSTOMERCLASS = col_character(),
##     COMBINED = col_character(),
##     TOTALCUSTOMERS = col_number(),
##     TOTALTHM = col_number(),
##     AVERAGETHM = col_double()
## )

##
## -- Column specification -----
## cols(
##     ZIPCODE = col_double(),
##     MONTH = col_double(),
##     YEAR = col_double(),
##     CUSTOMERCLASS = col_character(),
##     COMBINED = col_character(),
##     TOTALCUSTOMERS = col_number(),
##     TOTALTHM = col_number(),
##     AVERAGETHM = col_number()
## )
##
##
## -- Column specification -----
## cols(
##     ZIPCODE = col_double(),
##     MONTH = col_double(),
##     YEAR = col_double(),
##     CUSTOMERCLASS = col_character(),
##     COMBINED = col_character(),
##     TOTALCUSTOMERS = col_number(),
##     TOTALTHM = col_number(),
##     AVERAGETHM = col_number()
## )
##
##
## -- Column specification -----
## cols(

```

```

##  ZIPCODE = col_double(),
##  MONTH = col_double(),
##  YEAR = col_double(),
##  CUSTOMERCLASS = col_character(),
##  COMBINED = col_character(),
##  TOTALCUSTOMERS = col_number(),
##  TOTALTHM = col_number(),
##  AVERAGETHM = col_number()
## )

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_double()
## )

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )

##
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )

##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),

```



```

## CUSTOMERCLASS = col_character(),
## COMBINED = col_character(),
## TOTALCUSTOMERS = col_number(),
## TOTALTHM = col_number(),
## AVERAGETHM = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )
##
##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),

```

```

##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_double()
## )

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )

Quarters = 1:2

for (Quarter in Quarters) {

  Filename <- paste0("PGE/Gas/PGE_2021_Q",
    Quarter, "_", Type, "UsageByZip.csv")

  Temp <- read_csv(Filename)
  PGE_Gas <- rbind(PGE_Gas, Temp)
}

##
## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )
##
##

```

```

## -- Column specification -----
## cols(
##   ZIPCODE = col_double(),
##   MONTH = col_double(),
##   YEAR = col_double(),
##   CUSTOMERCLASS = col_character(),
##   COMBINED = col_character(),
##   TOTALCUSTOMERS = col_number(),
##   TOTALTHM = col_number(),
##   AVERAGETHM = col_number()
## )

PGE_Gas = data.frame(PGE_Gas)
rm(Temp)

names(PGE_Elec)[7] <- "TOTAL_KWH/THM"
names(PGE_Elec)[8] <- "AVERAGE_KWH/THM"

names(PGE_Gas)[7] <- "TOTAL_KWH/THM"
names(PGE_Gas)[8] <- "AVERAGE_KWH/THM"

finalData = rbind(PGE_Elec, PGE_Gas)

for (i in 1:nrow(finalData)) {

  if (startsWith(finalData$CUSTOMERCLASS[i],
    "Elec")) {
    finalData$TOTAL_kBTUs[i] = finalData$`TOTAL_KWH/THM`[i] *
      3.412

  } else if (startsWith(finalData$CUSTOMERCLASS[i],
    "Gas")) {
    finalData$TOTAL_kBTUs[i] = finalData$`TOTAL_KWH/THM`[i] *
      100
  }
}

ResidentialData = subset(finalData, finalData$CUSTOMERCLASS ==
  "Elec- Residential" | finalData$CUSTOMERCLASS ==
  "Gas- Residential")

CommercialData = subset(finalData, finalData$CUSTOMERCLASS ==
  "Elec- Commercial" | finalData$CUSTOMERCLASS ==
  "Gas- Commercial")

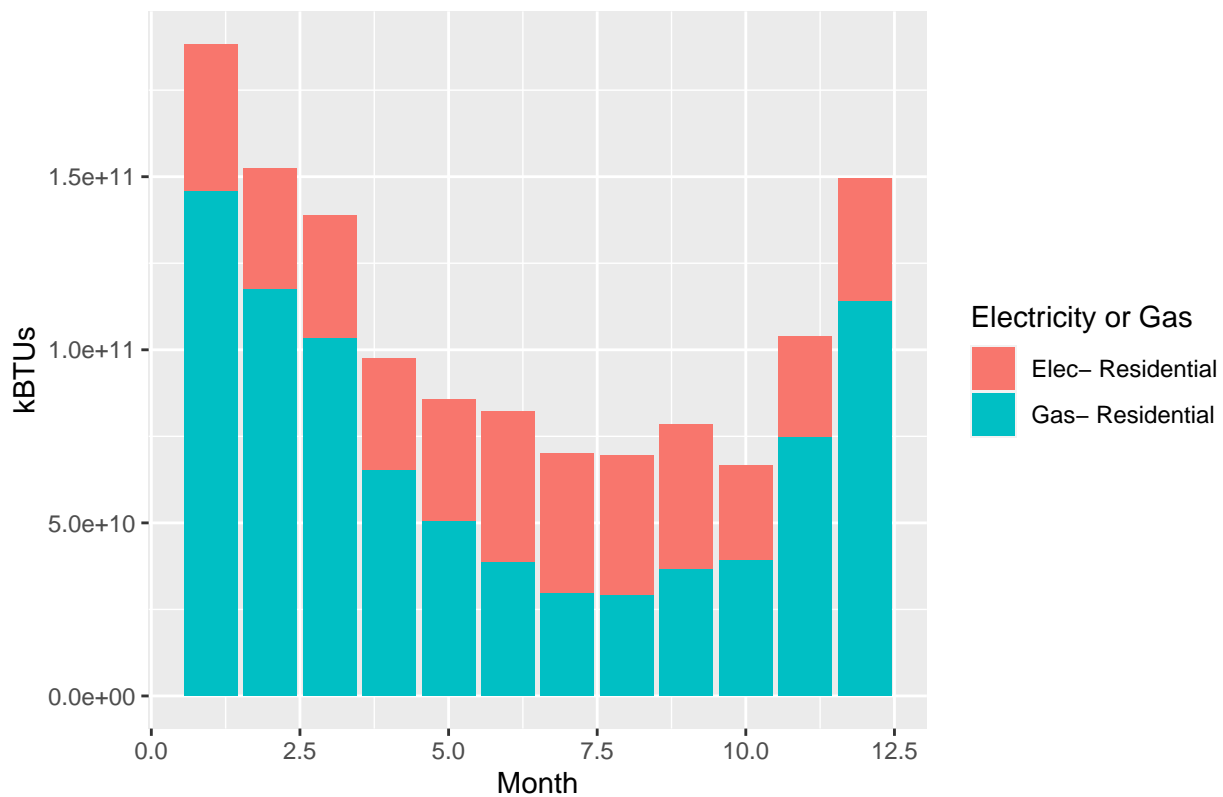
# Residential

newPlotOne = ggplot(data = ResidentialData) +
  geom_bar(aes(x = MONTH, y = TOTAL_kBTUs,
    fill = CUSTOMERCLASS), stat = "identity",
    position = "stack") + labs(x = "Month",
    y = "kBTUs", title = "PG&E Residential Data 2017-2020 + 2021 Q1 and Q2",
    fill = "Electricity or Gas")

newPlotOne

```

PG&E Residential Data 2017–2020 + 2021 Q1 and Q2

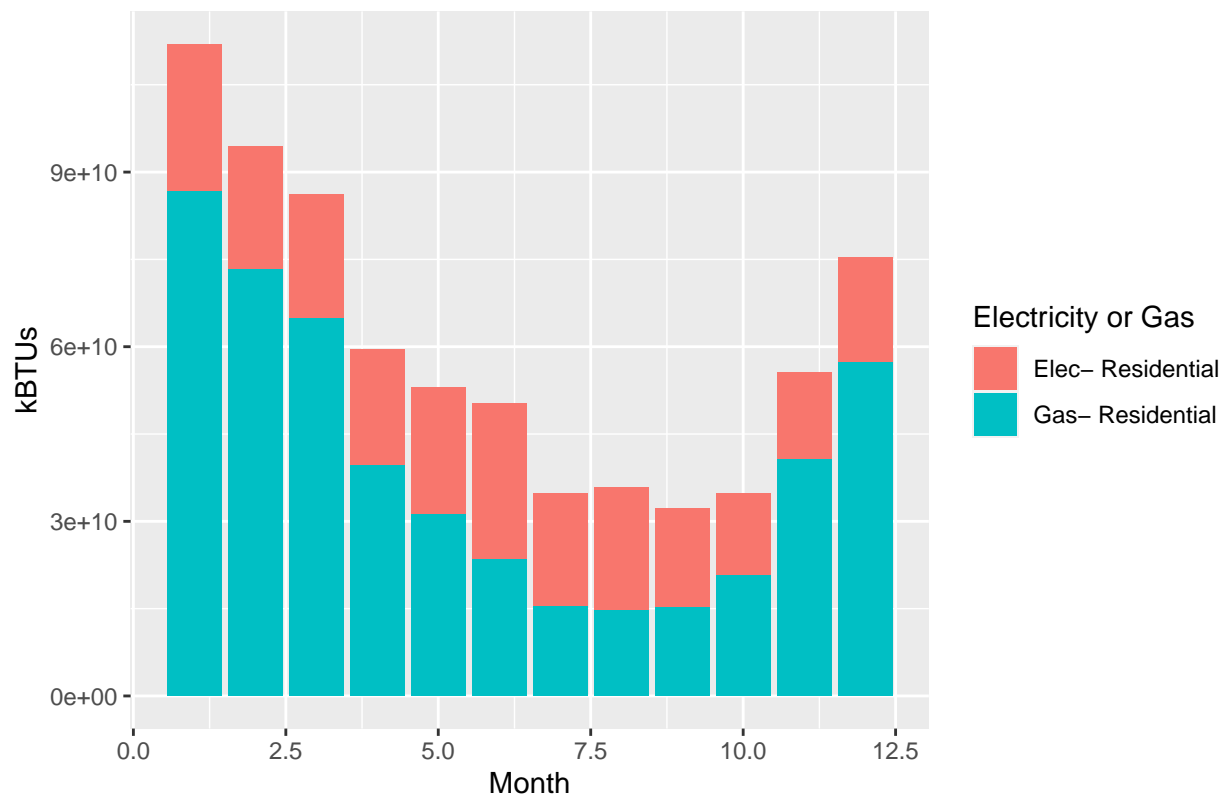


```
# Residential COVID Plot
COVIDResidential = subset(ResidentialData,
  ResidentialData$YEAR == "2019" | ResidentialData$YEAR ==
    "2020" | ResidentialData$YEAR ==
    "2021")

newPlotOneCOVID = ggplot(data = COVIDResidential) +
  geom_bar(aes(x = MONTH, y = TOTAL_kBTUs,
    fill = CUSTOMERCLASS), stat = "identity",
    position = "stack") + labs(x = "Month",
  y = "kBTUs", title = "PG&E Residential Data 2019-2020 + 2021 Q1 and Q2",
  fill = "Electricity or Gas")

newPlotOneCOVID
```

PG&E Residential Data 2019–2020 + 2021 Q1 and Q2

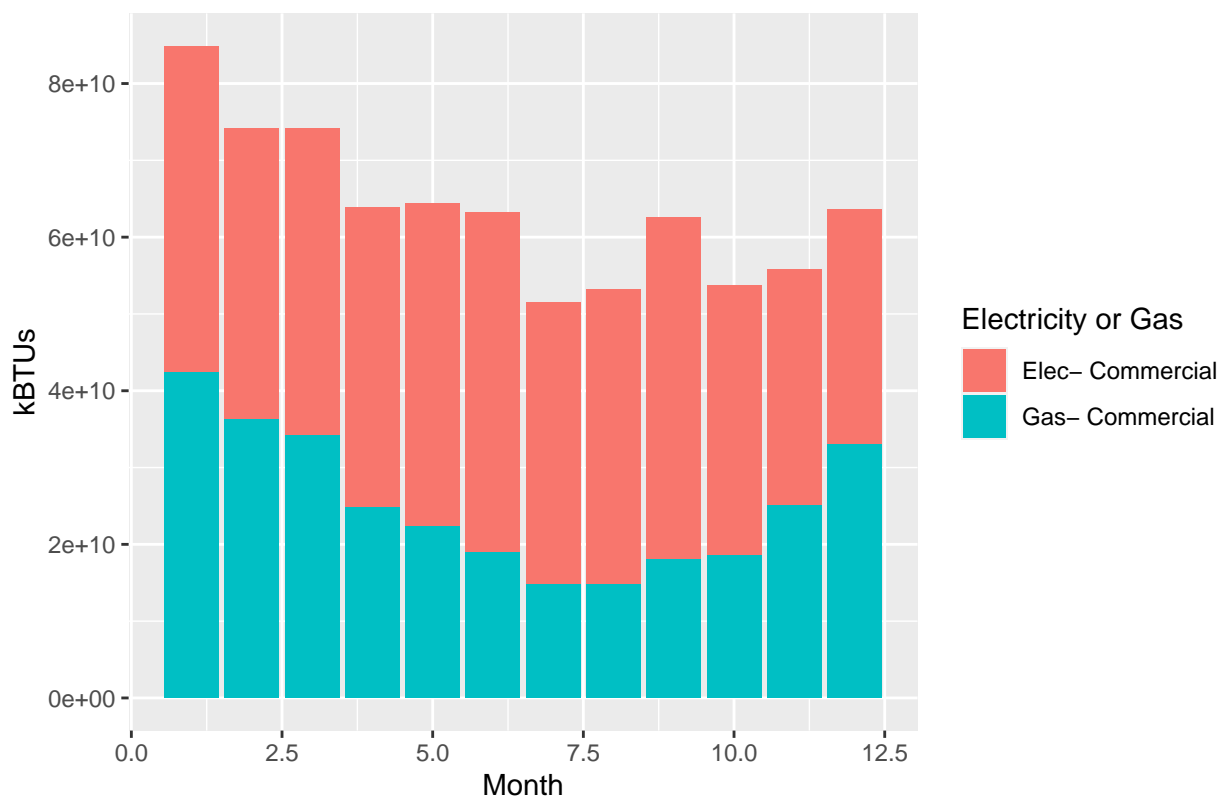


Commercial

```
newPlotTwo = ggplot(data = CommercialData) +
  geom_bar(aes(x = MONTH, y = TOTAL_kBTUs,
    fill = CUSTOMERCLASS), stat = "identity",
    position = "stack") + labs(x = "Month",
    y = "kBTUs", title = "PG&E Commercial Data 2017-2020 + 2021 Q1 and Q2",
    fill = "Electricity or Gas")
```

newPlotTwo

PG&E Commercial Data 2017–2020 + 2021 Q1 and Q2

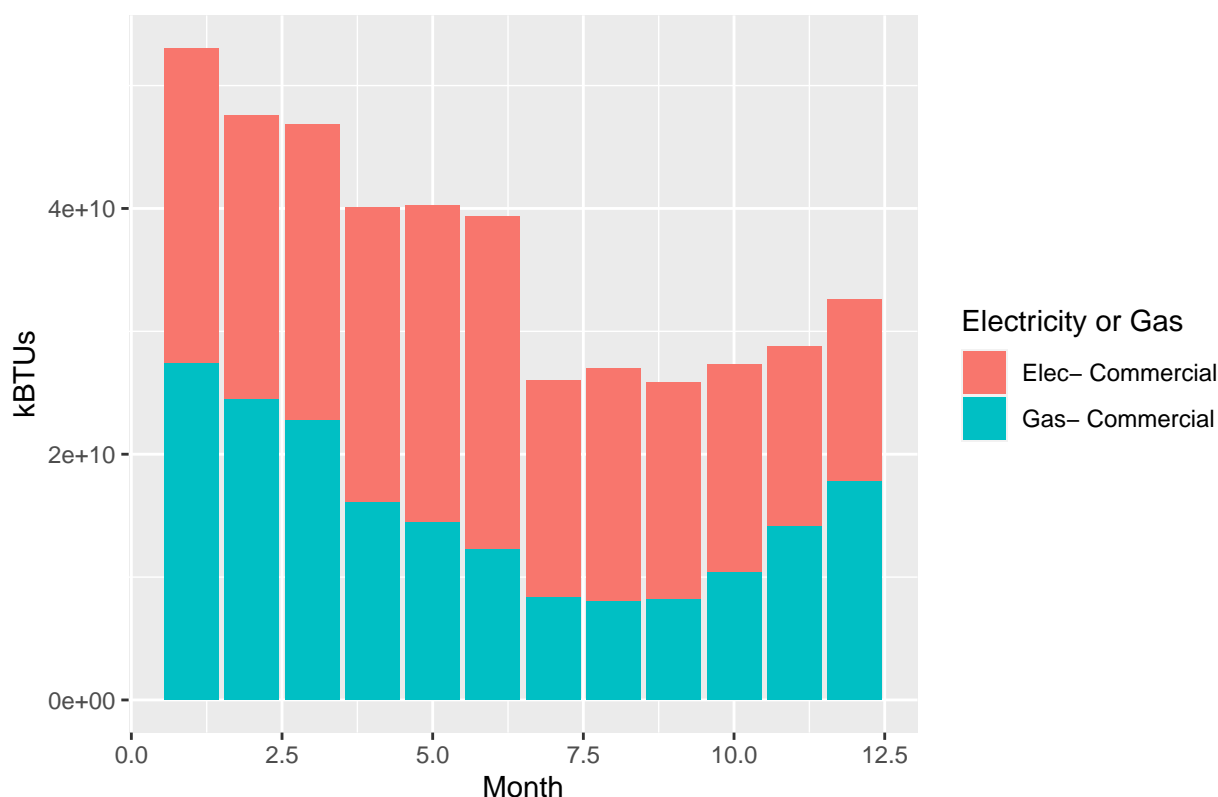


```
COVIDCommercial = subset(CommercialData,
  CommercialData$YEAR == "2019" | CommercialData$YEAR ==
    "2020" | CommercialData$YEAR == "2021")

newPlotTwoCOVID = ggplot(data = COVIDCommercial) +
  geom_bar(aes(x = MONTH, y = TOTAL_kBTUs,
    fill = CUSTOMERCLASS), stat = "identity",
    position = "stack") + labs(x = "Month",
    y = "kBTUs", title = "PG&E Commercial Data 2019-2020 + 2021 Q1 and Q2",
    fill = "Electricity or Gas")

newPlotTwoCOVID
```

PG&E Commercial Data 2019–2020 + 2021 Q1 and Q2



Summarizing Results & COVID Pandemic 2019 - 2021 Analysis

For both the Residential and Commercial Data between 2019 and 2021, it appears that the kBTU usages for Electricity and Gas are relatively lower across all months, give or take, in comparison to the Residential and Commercial Data between 2017 - 2021. To see this, in comparing the Residential and Commercial Data between 2019 and 2021 and 2017 and 2021, the range of the kBTU on the Y - Axis has shrunk. Ultimately, this suggests that more energy was used between 2017 and 2019. This makes sense, as because of the long-lasting impacts of COVID-19, many of the Residences and Commercial Institutions will have lost money/resources to pay for electricity and gas, so they may be more stingy and less willing to spend.

In terms of the key assumptions made in the analysis and/or caveats we should be aware of, note that the website (<https://pge-energydatarequest.com/>) states that the data contains “a minimum of 100 Residential Customers, a minimum of 15 Non-Residential Customers, with no single Non-Residential Customer in each sector accounting for more than 15% of the total consumption. If the aggregation standard is not met, the consumption will be combined with a neighboring ZIP code until the aggregation requirements are met.” Thus, it assumes that the if the aggregation standard is not met, the consumption combined with the neighboring ZIP code will have a negligible impact on drawing conclusions from the data. Moreover, we also assume that the data is well-representative of the population, and that the information is accurate, as the KWhs and Therms are accurately measured.