Assignment_1

2025-09-05

Task-1: Getting the dataset from web source

I downloaded the publicly accessible CNG stations' location data from the website of Department of Energy

Reference: Office of Energy Efficiency and Renewable Energy. (n.d.). Natural Gas Fueling Station Locations.US Department of Energy. Retrieved from https://afdc.energy.gov/fuels/natural-gas-locations#/find/nearest?fuel=CNG

Task-2: Importing the dataset dowloaded previously

```
library(readr)
CNG_stations <- read_csv(".\\CNG_stations.csv")
```

```
## Rows: 721 Columns: 39
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (30): Fuel Type Code, Station Name, Street Address, Intersection Directi...
## dbl  (6): Latitude, Longitude, ID, CNG Dispenser Num, CNG Total Compression ...
## lgl  (3): BD Blends, Hydrogen Is Retail, Restricted Access
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(CNG_stations)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Task-3: Descriptive analysis

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v purrr     1.0.2
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
#using quantitative variable
summary(CNG_stations$`CNG Total Compression Capacity`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##     2.0   300.0   745.0   874.7  1200.0  8000.0     285
```

```r
#state-wise number of cng stations
CNG_stations %>% group_by(`State`) %>% summarise(tot_station = n())
```

```
## # A tibble: 42 x 2
##    State tot_station
##    <chr>       <int>
##  1 AL              2
##  2 AR              8
##  3 AZ              9
##  4 CA            158
##  5 CO             16
##  6 CT              7
##  7 DE              1
##  8 FL             23
##  9 GA             17
## 10 IA              5
## # i 32 more rows
```

```r
#using categorical variable
summary(CNG_stations$`Facility Type`)
```

```
##    Length     Class      Mode 
##       721 character character
```

```r
#number of cng stations near different facilities
facility_stations <- CNG_stations %>% group_by(`Facility Type`) %>% summarise(station_no = n())
facility_stations
```

```
## # A tibble: 22 x 2
##    `Facility Type`   station_no
##    <chr>                  <int>
##  1 AIRPORT                   22
##  2 CARWASH                    1
##  3 COLLEGE_CAMPUS             4
##  4 CONVENIENCE_STORE        108
##  5 COOP                       2
##  6 FED_GOV                    1
##  7 FLEET_GARAGE              23
##  8 FUEL_RESELLER              4
##  9 GAS_STATION                8
## 10 GROCERY                    1
## # i 12 more rows
```

```
save(facility_stations,file=".\\facility_stations.csv")
```

Task-4: Transforming variable

```
is.logical(CNG_stations$`Restricted Access`)
```

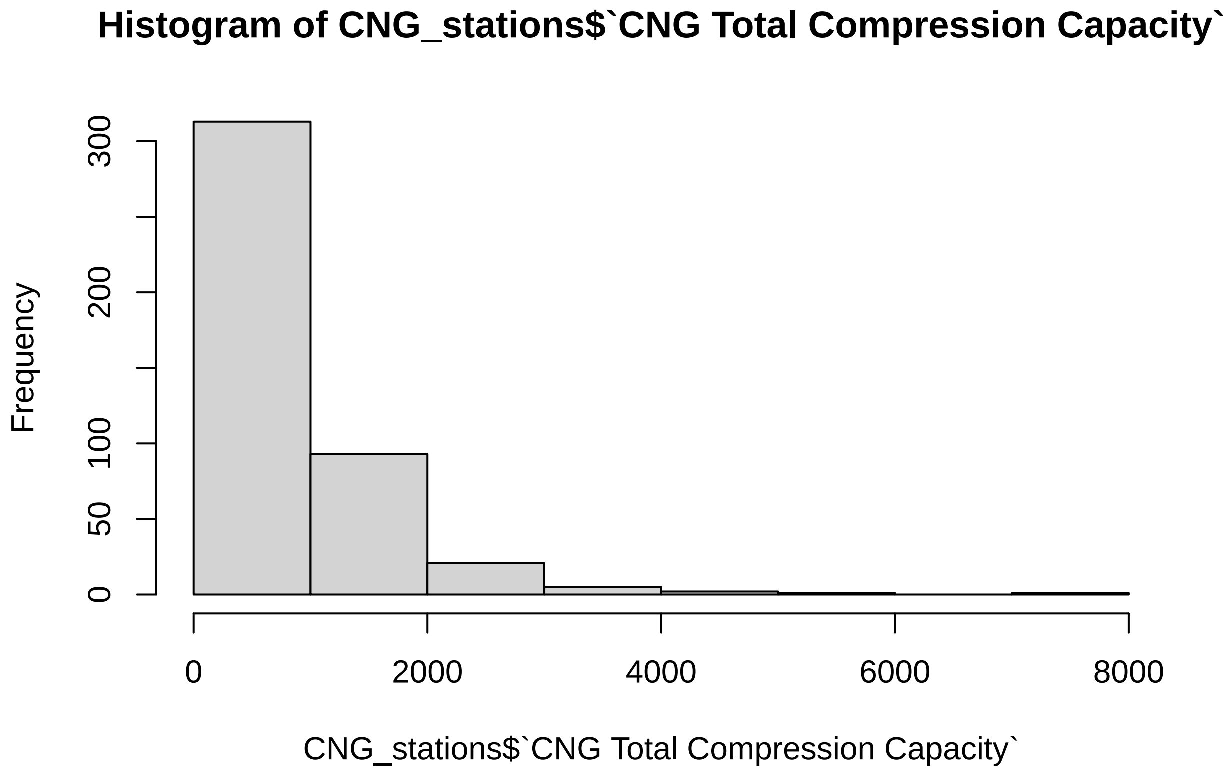
```
## [1] TRUE
```

```
#converting logical var to numeric
as.numeric(CNG_stations$`Restricted Access`)
```

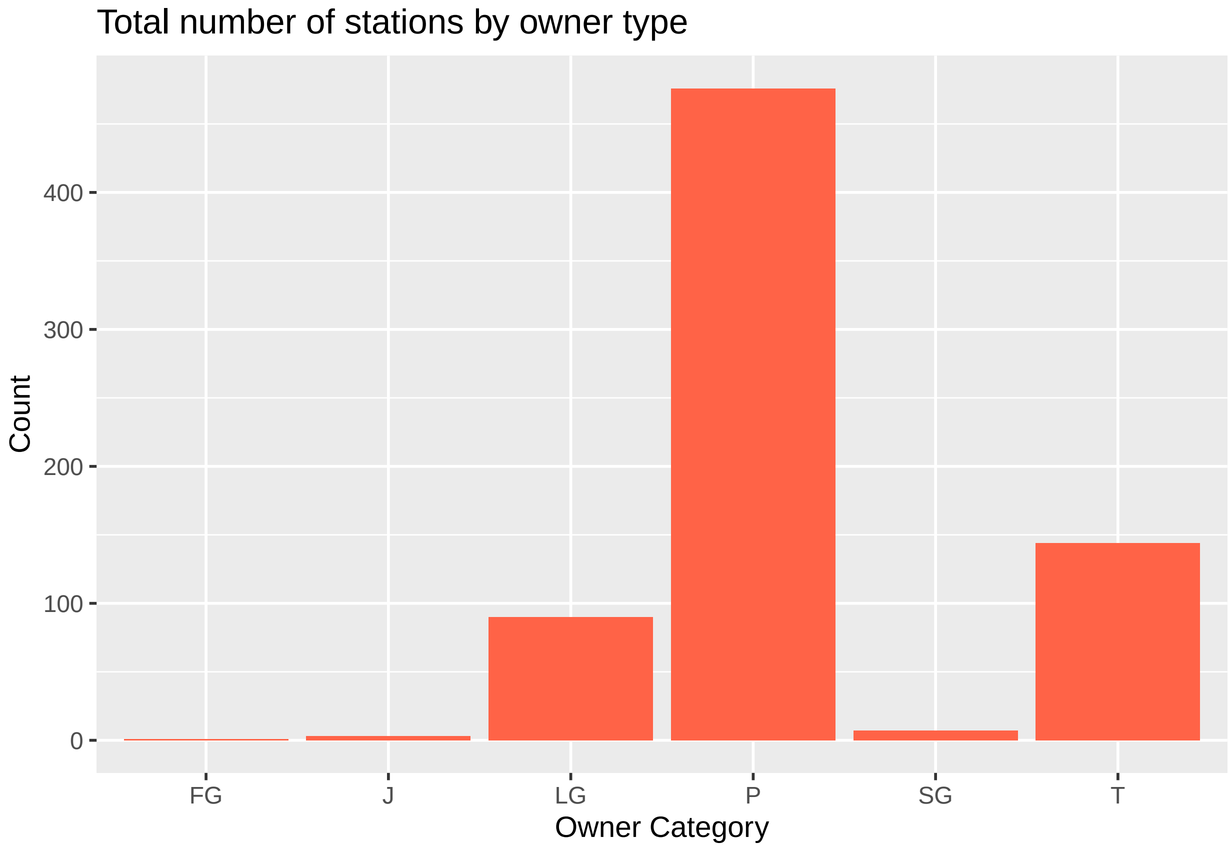
```
##   [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [38] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [75] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [112] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [149] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
## [186] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [223] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [260] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [297] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [334] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [371] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [408] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [445] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [482] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [519] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [556] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [593] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [630] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [667] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [704] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Task-5: Plotting

```
library(ggplot2)
#histogram
hist(CNG_stations$`CNG Total Compression Capacity`)
```

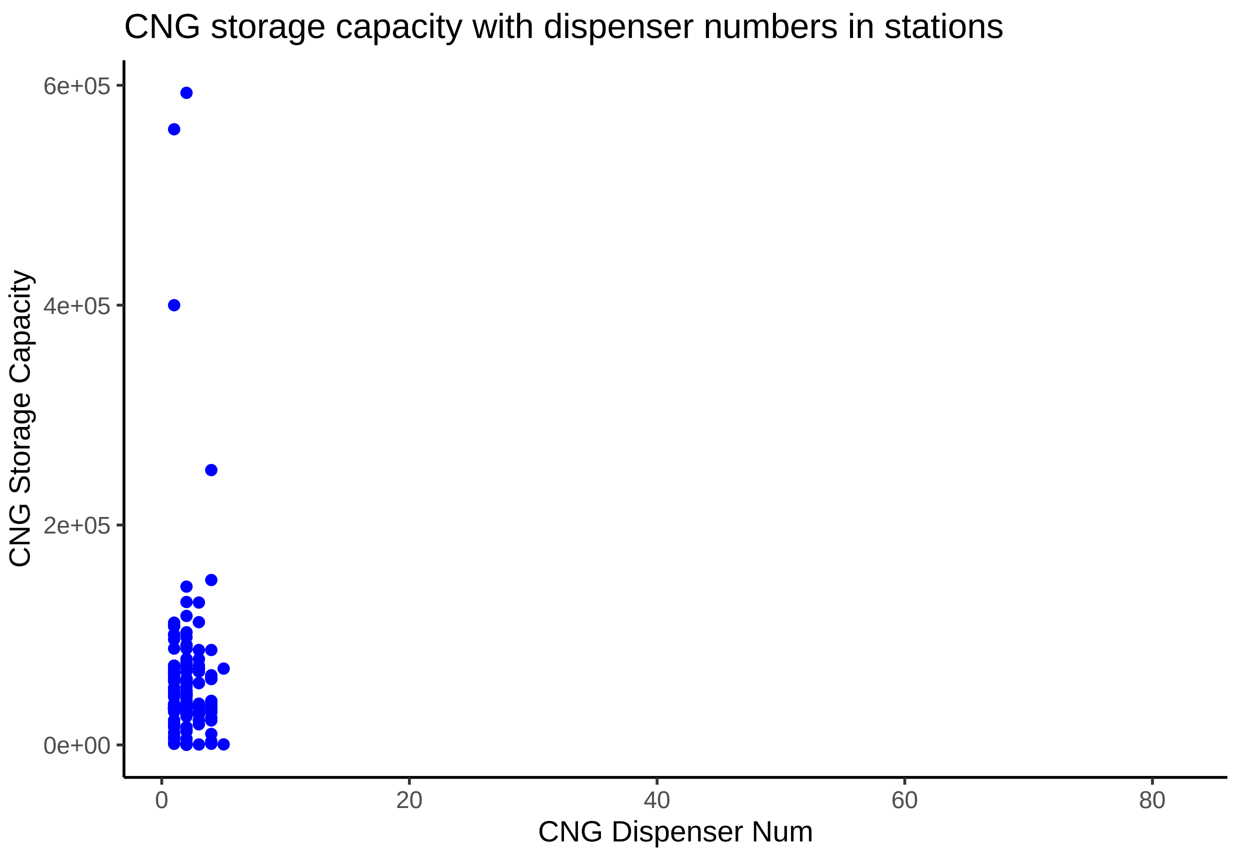


```r
#bar chart
ggplot(CNG_stations, aes(x = `Owner Type Code`)) +
  geom_bar(fill = "tomato") +
  labs(title = "Total number of stations by owner type", x = "Owner Category", y = "Count")
```



```r
#scatter plot
ggplot(CNG_stations, aes(x = `CNG Dispenser Num`, y = `CNG Storage Capacity`)) +
  geom_point(color = "blue", size = 1.5) +
  theme_classic()+
  labs(title = "CNG storage capacity with dispenser numbers in stations")
```

```
## Warning: Removed 489 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```


CNG storage capacity with dispenser numbers in stations
CNG Storage Capacity
6e+05
4e+05
2e+05
0e+00
0
20
40
60
80
CNG Dispenser Num