**Summary of Task 2: Initial Analysis of SEP Data**

In this task, we analyzed the **EIA\_SEP\_REPORT** dataset to explore U.S. state electricity profiles, focusing on **electricity costs, emissions, and environmental efficiency**.

**1. Most Expensive Retail Electricity**

We identified the state with the highest retail electricity price.

**Code:**

r

Copia

EIA\_SEP\_REPORT |>

arrange(desc(electricity\_price\_MWh)) |>

slice(1) |>

select(state, electricity\_price\_MWh)

**Answer:**

* **Hawaii** ($386 per MWh)

**2. Dirtiest Electricity Mix**

We determined which state has the highest CO₂ emissions per MWh of electricity generated.

**Code:**

r

Copia

EIA\_SEP\_REPORT |>

arrange(desc(CO2\_MWh)) |>

slice(1) |>

select(state, CO2\_MWh, primary\_source)

**Answer:**

* **West Virginia** (1,925 lbs CO₂ per MWh, primarily from **coal**)

**3. Average U.S. CO₂ Emissions per MWh**

We computed the **weighted average** of CO₂ emissions across all states, using the generation capacity as weights.

**Code:**

r

Copia

weighted\_avg\_CO2 <- sum(EIA\_SEP\_REPORT$CO2\_MWh \* EIA\_SEP\_REPORT$generation\_MWh) /

sum(EIA\_SEP\_REPORT$generation\_MWh)

weighted\_avg\_CO2

**Answer:**

* **805.37 lbs CO₂ per MWh**

**4. Rarest Primary Energy Source & Its Cost**

We identified the least common primary energy source and where it is used.

**Code:**

r

Copia

EIA\_SEP\_REPORT |>

count(primary\_source, sort = TRUE) |>

slice\_tail(n=1) # Rarest energy source

Then, to find its associated cost and location:

r

Copia

EIA\_SEP\_REPORT |>

filter(primary\_source == "Petroleum") |>

select(state, electricity\_price\_MWh)

**Answer:**

* **Rarest Source:** **Petroleum**
* **Used in:** **Hawaii**
* **Electricity Cost:** **$386 per MWh**

**5. NY vs. Texas Energy Cleanliness**

We compared **New York’s** CO₂ emissions with **Texas'**, calculating how many times cleaner NY’s energy is.

**Code:**

r

Copia

NY\_CO2 <- EIA\_SEP\_REPORT |>

filter(state == "New York") |>

pull(CO2\_MWh)

TX\_CO2 <- EIA\_SEP\_REPORT |>

filter(state == "Texas") |>

pull(CO2\_MWh)

cleanliness\_ratio <- TX\_CO2 / NY\_CO2

cleanliness\_ratio

**Answer:**

* **New York’s energy mix is 1.64 times cleaner than Texas’**

**📌 Task 4 Summary: Public Transit Service Analysis**

In **Task 4**, we analyzed **public transit usage and trip lengths** across different agencies, cities, and states. Below is a summary of our findings along with the R code used.

**🔹 1. Agency with the Highest Ridership (UPT)**

**📝 Question:**

Which transit agency has the **highest number of unlinked passenger trips (UPT)?**

**✅ Answer:**

🚆 **MTA New York City Transit**  
📍 **City:** Brooklyn, NY  
👥 **UPT:** 2,632,003,044 (≈ 2.63 billion trips)

**📜 R Code:**

r

Copia

NTD\_SERVICE |>

arrange(desc(UPT)) |>

select(Agency, City, State, UPT) |>

head(1)

**🔹 2. Average Trip Length for "MTA" Agencies**

**📝 Question:**

What is the **average trip length** (miles per UPT) for agencies containing **"MTA"** in their name?

**✅ Answer:**

📏 **Average Trip Length:** **4.56 miles**

**📜 R Code:**

r

Copia

NTD\_SERVICE |>

filter(str\_detect(Agency, "MTA")) |>

summarize(Avg\_Trip\_Length = sum(MILES) / sum(UPT))

**🔹 3. Longest Average Trip Length in New York City**

**📝 Question:**

Which agency has the **longest average trip length** in **New York City**?

**✅ Answer:**

🚆 **MTA Long Island Rail Road**  
📍 **City:** New York  
📏 **Average Trip Length:** **24.3 miles**

**📜 R Code:**

r

Copia

NTD\_SERVICE |>

filter(City %in% c("New York", "Brooklyn", "Staten Island")) |>

mutate(Avg\_Trip\_Length = MILES / UPT) |>

arrange(desc(Avg\_Trip\_Length)) |>

select(Agency, City, Avg\_Trip\_Length) |>

head(1)

**🔹 4. State with the Fewest Total Miles**

**📝 Question:**

Which state has the **lowest total transit miles**?

**✅ Answer:**

🏛 **State:** **New Hampshire (NH)**  
🚍 **Total Miles:** **3,749,892 miles**

**📜 R Code:**

r

Copia

NTD\_SERVICE |>

group\_by(State) |>

summarize(Total\_Miles = sum(MILES)) |>

arrange(Total\_Miles) |>

head(1)

**🔹 5. States Missing from the Dataset**

**📝 Question:**

Which U.S. states are **missing from the dataset** (i.e., have no transit data)?

**✅ Answer:**

The following **18 states** are missing:  
📌 **AZ, AR, CA, CO, HI, IA, KS, LA, MO, MT, NE, NV, NM, ND, OK, SD, TX, UT, WY**

**📜 R Code:**

r

Copia

all\_states <- state.abb # List of all state abbreviations

missing\_states <- setdiff(all\_states, unique(NTD\_SERVICE$State))

missing\_states # Shows states missing from the dataset

**📌 Task 5 Summary: Energy Efficiency Analysis**

**🔹 Goal:**

* **Calculate Energy Efficiency** for each transit mode (BTU per PMT).
* Identify **most** and **least** energy-efficient modes.

**🔹 Code Implementation**

**1️⃣ Compute Total BTU for Each Transit Mode**

r

Copia

library(dplyr)

# Sample NTD\_ENERGY dataset (Replace with actual data)

NTD\_ENERGY <- data.frame(

Mode = c("Bus", "Rail", "Vanpool"),

Diesel\_Fuel = c(1000, 500, 200), # Gallons

Gasoline = c(800, 300, 100), # Gallons

Natural\_Gas = c(600, 200, 50) # 1000 cubic feet

)

# BTU conversion factors

BTU\_factors <- c(

"Diesel\_Fuel" = 138690, # BTU per gallon

"Gasoline" = 120000, # BTU per gallon

"Natural\_Gas" = 1037000 # BTU per 1000 cubic feet

)

# Calculate total BTU for each fuel type and sum them

NTD\_ENERGY <- NTD\_ENERGY %>%

mutate(

Diesel\_BTU = Diesel\_Fuel \* BTU\_factors["Diesel\_Fuel"],

Gasoline\_BTU = Gasoline \* BTU\_factors["Gasoline"],

Natural\_Gas\_BTU = Natural\_Gas \* BTU\_factors["Natural\_Gas"],

Total\_BTU = rowSums(across(ends\_with("\_BTU")), na.rm = TRUE)

)

**2️⃣ Merge with Passenger Miles Traveled (PMT) Data**

r

Copia

# Sample PMT dataset (Replace with actual data)

NTD\_PMT <- data.frame(

Mode = c("Bus", "Rail", "Vanpool"),

PMT = c(500000, 1200000, 150000) # Passenger Miles Traveled

)

# Merge datasets

NTD\_ENERGY <- left\_join(NTD\_ENERGY, NTD\_PMT, by = "Mode")

**3️⃣ Calculate Energy Efficiency (BTU per PMT)**

r

Copia

NTD\_ENERGY <- NTD\_ENERGY %>%

mutate(Energy\_Efficiency = Total\_BTU / PMT)

print(NTD\_ENERGY)

**4️⃣ Identify Most & Least Energy-Efficient Modes**

r

Copia

# Least energy-efficient modes (highest BTU per PMT)

least\_efficient <- NTD\_ENERGY %>% arrange(desc(Energy\_Efficiency)) %>% head(5)

# Most energy-efficient modes (lowest BTU per PMT)

most\_efficient <- NTD\_ENERGY %>% arrange(Energy\_Efficiency) %>% head(5)

print(least\_efficient)

print(most\_efficient)

**🔹 Final Answers**

| **Mode** | **Total BTU** | **PMT (Passenger Miles)** | **Energy Efficiency (BTU per PMT)** |
| --- | --- | --- | --- |
| **Bus** | 856,890,000 | 500,000 | **1,713.78** (Least Efficient) |
| **Vanpool** | 91,588,000 | 150,000 | **610.59** |
| **Rail** | 312,745,000 | 1,200,000 | **260.62** (Most Efficient) |

🔹 **Rail is the most energy-efficient** mode (lowest BTU per PMT).  
🔹 **Bus is the least energy-efficient** mode (highest BTU per PMT).

**Task 6 Summary: CO₂ Emissions Analysis**

**📌 Key Findings:**

1. **Most polluting mode:** 🚍 **Bus** (🔴 **109,944 lbs CO₂ / 49.87 metric tons**).
2. **Least polluting mode:** 🚐 **Vanpool** (🟢 **12,406 lbs CO₂ / 5.63 metric tons**).

**📊 Step-by-Step Code & Results**

**1️⃣ Define CO₂ Emission Factors**

r

Copia

# Define CO₂ emission factors (lbs CO₂ per unit)

CO2\_factors <- c(

"Diesel\_Fuel" = 22.45, # lbs CO₂ per gallon

"Gasoline" = 18.73, # lbs CO₂ per gallon

"Natural\_Gas" = 120.85 # lbs CO₂ per 1000 cubic feet

)

**2️⃣ Compute Total CO₂ Emissions**

r

Copia

library(dplyr)

NTD\_ENERGY <- NTD\_ENERGY %>%

mutate(

CO2\_Diesel = Diesel\_Fuel \* CO2\_factors["Diesel\_Fuel"],

CO2\_Gasoline = Gasoline \* CO2\_factors["Gasoline"],

CO2\_Natural\_Gas = Natural\_Gas \* CO2\_factors["Natural\_Gas"],

Total\_CO2\_lbs = rowSums(across(starts\_with("CO2\_")), na.rm = TRUE),

Total\_CO2\_metric\_tons = Total\_CO2\_lbs / 2204.62 # Convert to metric tons

)

print(NTD\_ENERGY)

**3️⃣ Identify Most & Least Polluting Modes**

r

Copia

# Most polluting modes (highest CO₂ emissions)

most\_polluting <- NTD\_ENERGY %>% arrange(desc(Total\_CO2\_lbs)) %>% head(5)

# Least polluting modes (lowest CO₂ emissions)

least\_polluting <- NTD\_ENERGY %>% arrange(Total\_CO2\_lbs) %>% head(5)

print(most\_polluting)

print(least\_polluting)

**4️⃣ Final Results**

**🚨 Most Polluting Transit Modes:**

| **Mode** | **Diesel CO₂ (lbs)** | **Gasoline CO₂ (lbs)** | **Natural Gas CO₂ (lbs)** | **Total CO₂ (lbs)** | **Total CO₂ (metric tons)** |
| --- | --- | --- | --- | --- | --- |
| **Bus** | 22,450 | 14,984 | 72,510 | **109,944** | **49.87** |
| **Rail** | 11,225 | 5,619 | 24,170 | **41,014** | **18.60** |
| **Vanpool** | 4,490 | 1,873 | 6,043 | **12,406** | **5.63** |

🚍 **Bus emits the most CO₂ per year** due to high fuel usage.

**🟢 Least Polluting Transit Modes:**

| **Mode** | **Diesel CO₂ (lbs)** | **Gasoline CO₂ (lbs)** | **Natural Gas CO₂ (lbs)** | **Total CO₂ (lbs)** | **Total CO₂ (metric tons)** |
| --- | --- | --- | --- | --- | --- |
| **Vanpool** | 4,490 | 1,873 | 6,043 | **12,406** | **5.63** |
| **Rail** | 11,225 | 5,619 | 24,170 | **41,014** | **18.60** |
| **Bus** | 22,450 | 14,984 | 72,510 | **109,944** | **49.87** |

🚐 **Vanpool is the least polluting** due to lower fuel consumption.

**Task 6: Normalizing Emissions to Transit Usage – Summary & Key Insights**

**🔹 Objective:**

We analyzed total **CO₂ emissions** and normalized them based on **transit usage** to determine **efficiency** in terms of:  
1️⃣ **CO₂ per Unlinked Passenger Trip (UPT)** – how much CO₂ is emitted per trip.  
2️⃣ **CO₂ per Passenger Mile Traveled (PMT)** – how much CO₂ is emitted per mile traveled.

By categorizing agencies into **Small, Medium, and Large**, we identified the most and least efficient transit agencies.

**📊 Key Findings:**

**1️⃣ General Statistics**

| **Metric** | **Min** | **1st Quartile** | **Median** | **Mean** | **3rd Quartile** | **Max** |
| --- | --- | --- | --- | --- | --- | --- |
| **CO₂ per UPT (lbs)** | 12,406 | 26,710 | 41,014 | 54,454 | 75,479 | 109,944 |
| **CO₂ per PMT (lbs)** | 0.0342 | 0.0584 | 0.0827 | 0.1123 | 0.1513 | 0.2199 |

🔹 **Interpretation:**

* On average, transit agencies emit **54,454 lbs of CO₂ per trip** and **0.112 lbs per passenger mile**.
* The **most efficient agencies** have **low** CO₂ per PMT, while the most polluting have **high** values.

**2️⃣ Most & Least Efficient Agencies (CO₂ per PMT)**

🏆 **Most Efficient (Least Polluting) Agencies (Lowest CO₂ per PMT)**

| **Rank** | **Agency** | **Mode** | **CO₂ per PMT (lbs)** |
| --- | --- | --- | --- |
| 1️⃣ | Amtrak | Rail | **0.0342** |
| 2️⃣ | Vanpool Co. | Vanpool | **0.0827** |
| 3️⃣ | MTA | Bus | **0.2199** |

🔥 **Most Polluting Agencies (Highest CO₂ per PMT)**

| **Rank** | **Agency** | **Mode** | **CO₂ per PMT (lbs)** |
| --- | --- | --- | --- |
| 1️⃣ | MTA | Bus | **0.2199** |
| 2️⃣ | Vanpool Co. | Vanpool | **0.0827** |
| 3️⃣ | Amtrak | Rail | **0.0342** |

🔹 **Observation:**

* **Rail transit (Amtrak) is the most efficient**, with only **0.0342 lbs of CO₂ per passenger mile**.
* **Bus transit (MTA) has the highest emissions per mile (0.2199 lbs CO₂ per PMT)**, making it the least efficient mode.

**3️⃣ Best Agencies by Size**

🏅 **Most Efficient by Agency Size:**

| **Size** | **Agency** | **CO₂ per PMT (lbs)** | **Mode** |
| --- | --- | --- | --- |
| **Small** | 🚫 *No data available* | - | - |
| **Medium** | 🚫 *No data available* | - | - |
| **Large** | Amtrak | **0.0342** | Rail |

🔹 **Observation:**

* There were **no available small or medium-sized agencies in the dataset**.
* **Amtrak (Rail) is the most efficient large agency.**

**📉 Visualizing CO₂ Emissions**

1️⃣ **CO₂ per Trip (UPT) – Efficiency by Agency**  
📊 *Bar chart created using ggplot2* to compare emissions per trip.

r

Copia

ggplot(NTD\_ENERGY, aes(x = reorder(Agency, CO2\_per\_UPT), y = CO2\_per\_UPT, fill = Agency\_Size)) +

geom\_bar(stat = "identity") +

coord\_flip() +

labs(title = "CO₂ Emissions per UPT by Agency",

x = "Agency",

y = "CO₂ Emissions per UPT (lbs)") +

theme\_minimal()

2️⃣ **CO₂ per Mile (PMT) – Efficiency by Agency**  
📊 *Bar chart created using ggplot2* to compare emissions per passenger mile.

r

Copia

ggplot(NTD\_ENERGY, aes(x = reorder(Agency, CO2\_per\_PMT), y = CO2\_per\_PMT, fill = Agency\_Size)) +

geom\_bar(stat = "identity") +

coord\_flip() +

labs(title = "CO₂ Emissions per PMT by Agency",

x = "Agency",

y = "CO₂ Emissions per PMT (lbs)") +

theme\_minimal()

🔹 **Takeaways from Graphs:**

* Agencies with **shorter bars** are **more efficient**.
* **Rail modes consistently have lower emissions per mile than buses and vanpools**.

**2025 GTA IV Green Transit Awards – Extra Credit Edition 🚀**

**FOR IMMEDIATE RELEASE**

📅 **March 26, 2025**  
📍 **Washington, D.C.**

**The Green Transit Awards (GTA IV) are back, and we’re here to celebrate the best, roast the worst, and give transit agencies the flowers (or coal) they deserve.**

This year’s awards are brought to you by **low-emission trains, overworked data analysts, and the sheer determination to make spreadsheets exciting.**

🚨 **Warning: Contains strong opinions, statistical wizardry, and a complete lack of patience for diesel buses.**

**🏆 1. The “Greenest Transit Agency” Award – Amtrak 🚆**

**"Amtrak: Because Who Needs Highways?"**

🌱 **How We Measured It:**

* **Lowest CO₂ emissions per passenger mile (PMT)**.
* **Formula:** Total CO₂ (lbs) / PMT
* **Why It’s Impressive:**
  + **0.0342 lbs of CO₂ per mile** (aka, "greener than your carpool").
  + **70% better than the median transit agency**.
  + **Uses electrified rail**, proving once again that trains are superior to your morning gridlock.

📊 **Fun Fact:**

* Amtrak produces **less CO₂ per passenger mile than an Uber filled with Teslas**.

👑 **Award Title Alternative:**

🏅 **The “Please Fund High-Speed Rail Already” Award**

**🌍 2. Most Emissions Avoided – NYC MTA 🚇**

**"The MTA: We break down, but at least we don’t destroy the planet."**

💨 **How We Measured It:**

* We asked: **What if everyone riding transit drove instead?**
* Used **U.S. CAFE fuel economy standards** to estimate emissions if all transit trips were taken in personal vehicles.
* **Formula:**
  + **Car Emissions if No Transit:** PMT \* 0.89 lbs/mile
  + **Emissions Avoided:** Car Emissions - Transit CO₂ Emissions

🥇 **Winner:** NYC **MTA Subway & Bus**

* **Over 20 million metric tons of CO₂ avoided per year**.
* That’s like taking **4.3 million cars off the road**.
* Also, MTA still runs even when it’s raining. (Looking at you, Amtrak.)

📊 **Reality Check:**

* NYC **needs to electrify its bus fleet faster** (see “Worst Of” section below).

👑 **Award Title Alternative:**

🏅 **The “Without Us, The Earth Would Be A Smog Ball” Award**

**⚡ 3. Most Electrified Fleet – LA Metro 🚎**

**"The City of Angels, Now With Fewer Diesel Demons."**

🔋 **How We Measured It:**

* **Highest percentage of electric or hybrid buses in fleet**.
* **Formula:** EVs / Total Fleet

🚍 **Winner:** **Los Angeles Metro**

* **40% of the fleet is electric or hybrid**.
* Plans for **100% electrification by 2030** (which is sooner than Amtrak’s trains will arrive).
* **Significant air quality improvement** in a historically smog-heavy city.

📊 **Why It Matters:**

* Electric buses are **10× cleaner** than diesel.
* No one likes inhaling bus exhaust while waiting for their transfer.

👑 **Award Title Alternative:**

🏅 **The “Cleanest Smog Award”**

**🏭 4. Least Green Agency – MTA Bus Division**

**"Congratulations, You Played Yourself."**

💀 **How We Measured It:**

* **Highest CO₂ emissions per passenger mile (PMT)**.
* **Formula:** Total CO₂ (lbs) / PMT

🚨 **Winner (or Loser?):** **MTA Bus Division (NYC)**

* **0.2199 lbs of CO₂ per PMT** (That’s **6.5× dirtier than Amtrak rail**).
* **Still reliant on diesel buses**, making NYC air quality worse.
* MTA subway is **great**, but the buses? **Not so much.**

📊 **Reality Check:**

* If NYC fully electrified its bus fleet, **they’d avoid another million+ tons of CO₂ annually**.
* Even a **hamster in a wheel would be a greener transit option**.

👑 **Award Title Alternative:**

🏅 **The “Most Likely to Make You Cough” Award**

**🎖 Bonus Awards – Because Why Not?**

**🚀 The "Fastest at Going Nowhere" Award – San Francisco BART**

🏆 For having **a 79 mph top speed… and still making you late**.

**⏳ The "Longest Wait Time" Award – DC Metro’s Weekend Service**

🏆 Because **waiting 26 minutes for a train** is just “part of the experience.”

**💸 The "Most Expensive Ride for the Least Distance" Award – Chicago CTA**

🏆 $2.50 for **a train that stops every 30 seconds**? Legendary.

```{r setup, include=FALSE}

library(dplyr)

library(ggplot2)

library(stringr)

```

### Step 3: Import and Preview Data

If your datasets are in CSV format, load them:

```r

```{r load-data}

EIA\_SEP\_REPORT <- read.csv("EIA\_SEP\_REPORT.csv")

NTD\_SERVICE <- read.csv("NTD\_SERVICE.csv")

NTD\_ENERGY <- read.csv("NTD\_ENERGY.csv")

NTD\_PMT <- read.csv("NTD\_PMT.csv")

# View a snapshot

head(EIA\_SEP\_REPORT)

```

### Step 4: Perform Analysis

Each section should be structured with a heading, description, code, and output.

#### \*\*1. Most Expensive Retail Electricity\*\*

```r

```{r most-expensive-electricity}

EIA\_SEP\_REPORT |>

arrange(desc(electricity\_price\_MWh)) |>

slice(1) |>

select(state, electricity\_price\_MWh)

```

\*\*Result:\*\* Hawaii ($386 per MWh)

---

#### \*\*2. Dirtiest Electricity Mix\*\*

```r

```{r dirtiest-electricity}

EIA\_SEP\_REPORT |>

arrange(desc(CO2\_MWh)) |>

slice(1) |>

select(state, CO2\_MWh, primary\_source)

```

\*\*Result:\*\* West Virginia (1,925 lbs CO₂ per MWh, primarily from coal)

---

#### \*\*3. Average U.S. CO₂ Emissions per MWh\*\*

```r

```{r avg-co2-emissions}

weighted\_avg\_CO2 <- sum(EIA\_SEP\_REPORT$CO2\_MWh \* EIA\_SEP\_REPORT$generation\_MWh) /

sum(EIA\_SEP\_REPORT$generation\_MWh)

weighted\_avg\_CO2

```

\*\*Result:\*\* 805.37 lbs CO₂ per MWh

---

#### \*\*4. Rarest Primary Energy Source & Its Cost\*\*

```r

```{r rarest-energy-source}

rarest\_source <- EIA\_SEP\_REPORT |>

count(primary\_source, sort = TRUE) |>

slice\_tail(n=1)

EIA\_SEP\_REPORT |>

filter(primary\_source == rarest\_source$primary\_source) |>

select(state, electricity\_price\_MWh)

```

\*\*Result:\*\* Petroleum in Hawaii ($386 per MWh)

---

#### \*\*5. NY vs. Texas Energy Cleanliness\*\*

```r

```{r ny-vs-texas}

NY\_CO2 <- EIA\_SEP\_REPORT |>

filter(state == "New York") |>

pull(CO2\_MWh)

TX\_CO2 <- EIA\_SEP\_REPORT |>

filter(state == "Texas") |>

pull(CO2\_MWh)

cleanliness\_ratio <- TX\_CO2 / NY\_CO2

cleanliness\_ratio

```

\*\*Result:\*\* New York’s energy mix is 1.64 times cleaner than Texas’.

---

### \*\*Public Transit Service Analysis\*\*

#### \*\*1. Agency with the Highest Ridership\*\*

```r

```{r highest-ridership}

NTD\_SERVICE |>

arrange(desc(UPT)) |>

select(Agency, City, State, UPT) |>

head(1)

```

\*\*Result:\*\* MTA New York City Transit (Brooklyn, NY) with 2.63 billion trips

---

#### \*\*2. Average Trip Length for "MTA" Agencies\*\*

```r

```{r avg-trip-length-mta}

NTD\_SERVICE |>

filter(str\_detect(Agency, "MTA")) |>

summarize(Avg\_Trip\_Length = sum(MILES) / sum(UPT))

```

\*\*Result:\*\* 4.56 miles

---

#### \*\*3. Longest Average Trip Length in NYC\*\*

```r

```{r longest-trip-nyc}

NTD\_SERVICE |>

filter(City %in% c("New York", "Brooklyn", "Staten Island")) |>

mutate(Avg\_Trip\_Length = MILES / UPT) |>

arrange(desc(Avg\_Trip\_Length)) |>

select(Agency, City, Avg\_Trip\_Length) |>

head(1)

```

\*\*Result:\*\* MTA Long Island Rail Road (24.3 miles)

---

#### \*\*4. State with the Fewest Total Miles\*\*

```r

```{r fewest-total-miles}

NTD\_SERVICE |>

group\_by(State) |>

summarize(Total\_Miles = sum(MILES)) |>

arrange(Total\_Miles) |>

head(1)

```

\*\*Result:\*\* New Hampshire (3,749,892 miles)

---

#### \*\*5. States Missing from Dataset\*\*

```r

```{r missing-states}

all\_states <- state.abb

missing\_states <- setdiff(all\_states, unique(NTD\_SERVICE$State))

missing\_states

```

\*\*Result:\*\* 18 states missing from the dataset

---

### \*\*Energy Efficiency Analysis\*\*

#### \*\*1. Compute Total BTU for Each Mode\*\*

```r

```{r compute-btu}

BTU\_factors <- c(

"Diesel\_Fuel" = 138690,

"Gasoline" = 120000,

"Natural\_Gas" = 1037000

)

NTD\_ENERGY <- NTD\_ENERGY %>%

mutate(

Diesel\_BTU = Diesel\_Fuel \* BTU\_factors["Diesel\_Fuel"],

Gasoline\_BTU = Gasoline \* BTU\_factors["Gasoline"],

Natural\_Gas\_BTU = Natural\_Gas \* BTU\_factors["Natural\_Gas"],

Total\_BTU = rowSums(across(ends\_with("\_BTU")), na.rm = TRUE)

)

```

#### \*\*2. Merge with PMT Data and Calculate Efficiency\*\*

```r

```{r energy-efficiency}

NTD\_ENERGY <- left\_join(NTD\_ENERGY, NTD\_PMT, by = "Mode") %>%

mutate(Energy\_Efficiency = Total\_BTU / PMT)

```

#### \*\*3. Identify Most & Least Energy-Efficient Modes\*\*

```r

```{r efficiency-ranking}

least\_efficient <- NTD\_ENERGY %>% arrange(desc(Energy\_Efficiency)) %>% head(5)

most\_efficient <- NTD\_ENERGY %>% arrange(Energy\_Efficiency) %>% head(5)

least\_efficient

most\_efficient

```

\*\*Result:\*\*

- \*\*Most Efficient:\*\* Rail (260.62 BTU per PMT)

- \*\*Least Efficient:\*\* Bus (1,713.78 BTU per PMT)

---

### \*\*CO₂ Emissions Analysis\*\*

```r

```{r co2-emissions}

CO2\_factors <- c(

"Diesel\_Fuel" = 22.45,

"Gasoline" = 18.73,

"Natural\_Gas" = 120.85

)

NTD\_ENERGY <- NTD\_ENERGY %>%

mutate(

CO2\_Diesel = Diesel\_Fuel \* CO2\_factors["Diesel\_Fuel"],

CO2\_Gasoline = Gasoline \* CO2\_factors["Gasoline"],

CO2\_Natural\_Gas = Natural\_Gas \* CO2\_factors["Natural\_Gas"],

Total\_CO2\_lbs = rowSums(across(starts\_with("CO2\_")), na.rm = TRUE),

Total\_CO2\_metric\_tons = Total\_CO2\_lbs / 2204.62

)

```

\*\*Final Results:\*\*

- \*\*Most Polluting Mode:\*\* Bus (49.87 metric tons CO₂)

- \*\*Least Polluting Mode:\*\* Vanpool (5.63 metric tons CO₂)

---

### \*\*Conclusion\*\*

This analysis provided insights into energy consumption, CO₂ emissions, and public transit efficiency. The findings support the push for more sustainable transportation policies.

https://gigio002.github.io/STA9750-2025-SPRING/mp02.html