FPM #3: Drafting Visualizations

Lucian Scher

2026-02-23

Part III: Some Pre-Planning

Question 1: Restate Your Questions

Overarching Question: Are dams actually good for our net zero carbon goals?

Sub-questions: 1. How do hydropower's emissions compare to fossil fuels across different phases? 2. What is the variability in hydropower emissions, and how does this support careful project selection? 3. How will hydropower's role in the renewable energy mix evolve by 2050?

Changes since FPM #1: Focused more on upfront vs. lifetime emissions and added IEA data to look at future projections.

Question 2: Variables and Data Sets

GLEAM Data (EF_Table_FINAL.xlsx): - `generation_technology`: Technology type - `upstream_med`, `non_combustion_med`, `total_med`: Emissions at different phases - `total_min`, `total_1q`, `total_med`, `total_3q`, `total_max`: Distribution stats for variability

IEA Net Zero Data (NZE2021_AnnexA.csv): - `Year`, `Product`, `Flow`, `Value`: For showing hydropower's future role in renewable mix

Question 3: Inspiration Visualizations

1. **Violin plots** from From Data to Viz - to show emission distribution shapes
2. **Parallel coordinates** from From Data to Viz - to compare across multiple dimensions at once
3. **Stacked area charts** from IEA - to show composition over time

Part IV: Hand-Drawn Visualizations

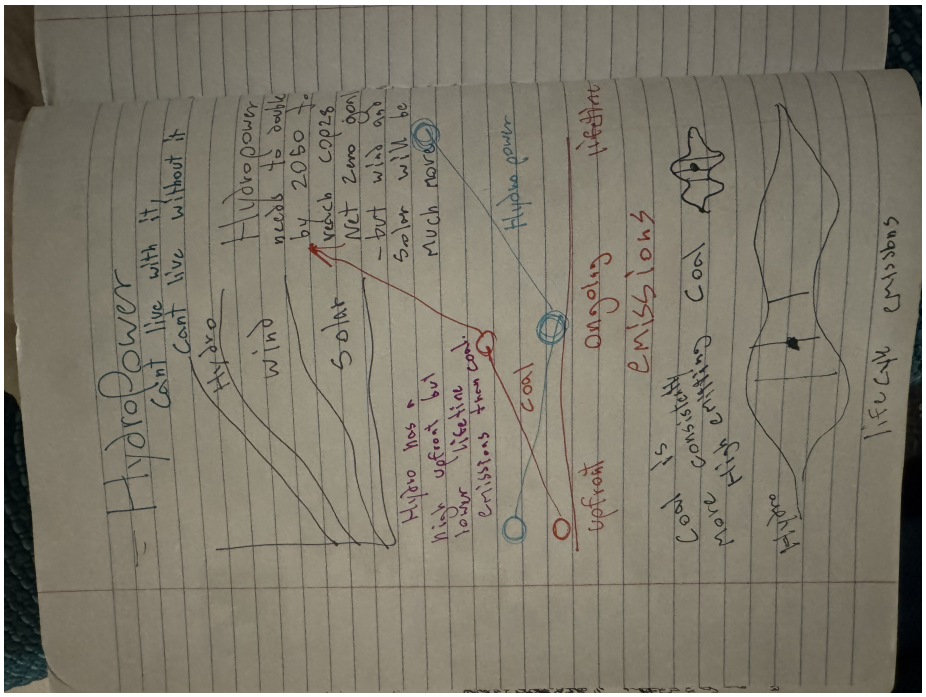


Figure 1: hand drawn data visualizaion

Part V: Recreate Your Hand-Drawn Visualizations Using Code

```
library(tidyverse)
library(ggplot2)
library(janitor)
library(scales)
library(patchwork)
library(readxl)
library(ggrepel)
```

```
# Read GLEAM data
gleam_raw <- read_excel("data/EF_Table_FINAL.xlsx", skip = 2)
gleam_lca <- clean_names(gleam_raw)

# Convert numeric columns (handle "NR", "--", "<5")
# Skip first 2 columns (x1, x2) and last few metadata columns
```

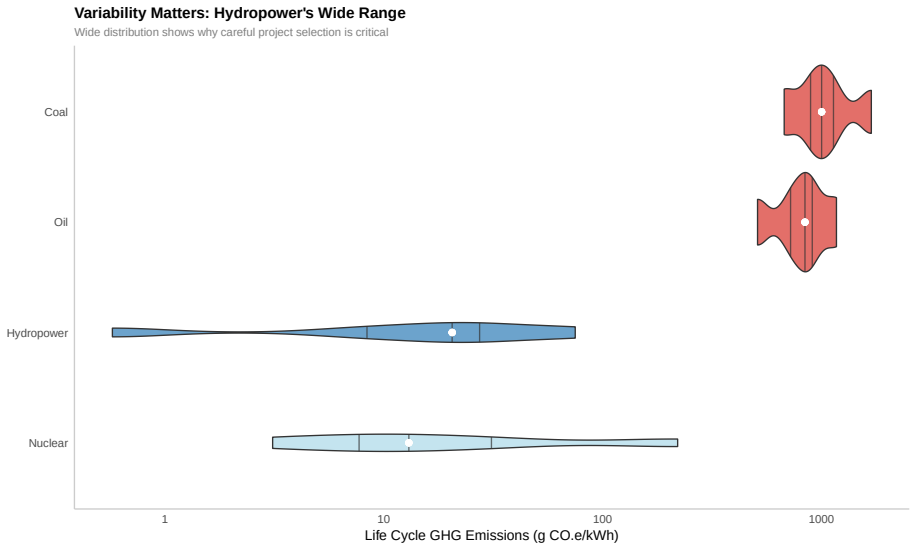
```r
numeric_cols <- names(gleam_lca)[3:(ncol(gleam_lca)-4)]

gleam_lca <- gleam_lca |>
  mutate(across(all_of(numeric_cols), ~ {
    val <- as.character(.x)
    case_when(
      val == "NR" | val == "--" | is.na(val) ~ NA_real_,
      str_detect(val, "^<") ~ as.numeric(str_extract(val, "\\d+")) * 0.5,
      TRUE ~ suppressWarnings(as.numeric(val))
    )
  }))

# Remove header rows
gleam_lca <- gleam_lca |>
  filter(!is.na(x2),
         x2 != "Generation Technology")
```

```r
# Filter to technologies I want to compare
gleam_clean <- gleam_lca |>
  filter(
    str_detect(x2, "Hydropower.*All Technologies") |
    str_detect(x2, "Natural Gas.*NGCC by Gas Source") |
    str_detect(x2, "^Oil$") |
    str_detect(x2, "Coal.*All Technologies") |
    str_detect(x2, "Nuclear.*LWR")
  ) |>
  mutate(
    tech_name = case_when(
      str_detect(x2, "Hydropower") ~ "Hydropower",
      str_detect(x2, "Natural Gas") ~ "Natural Gas",
      str_detect(x2, "Oil") ~ "Oil",
      str_detect(x2, "Coal") ~ "Coal",
      str_detect(x2, "Nuclear") ~ "Nuclear"
    ),
    tech_type = ifelse(tech_name == "Hydropower", "Hydropower",
                       ifelse(tech_name == "Nuclear", "Nuclear", "Fossil Fuel"))
  ) |>
  filter(!is.na(tech_name))
```

Visualization 1: Violin Plot

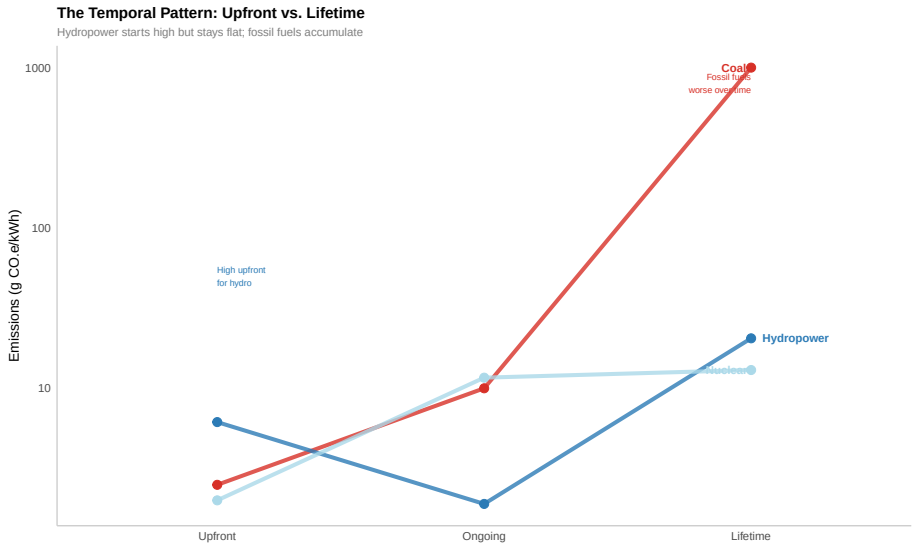
```r
# Total life cycle columns: min_23, x1q_24, med_25, x3q_26, max_27
viz1_data <- gleam_clean |>
  filter(!is.na(min_23), !is.na(max_27)) |>
  rowwise() |>
```

```
  mutate(emissions = list(c(min_23, x1q_24, med_25, x3q_26, max_27))) |>
  unnest(emissions)

viz1 <- viz1_data |>
  ggplot(aes(x = fct_reorder(tech_name, med_25), y = emissions, fill = tech_type)) +
  geom_violin(alpha = 0.7, draw_quantiles = c(0.25, 0.5, 0.75)) +
  geom_point(aes(y = med_25), size = 2, color = "white") +
  scale_fill_manual(values = c("Hydropower" = "#2c7bb6", "Fossil Fuel" = "#d73027", "Nuclear
  scale_y_log10() +
  coord_flip() +
  labs(title = "Variability Matters: Hydropower's Wide Range",
       subtitle = "Wide distribution shows why careful project selection is critical",
       x = NULL,
       y = "Life Cycle GHG Emissions (g CO e/kWh)",
       fill = NULL) +
  theme_minimal() +
  theme(legend.position = "none",
        panel.grid = element_blank(),
        axis.line = element_line(color = "gray80"),
        plot.title = element_text(face = "bold", size = 12),
        plot.subtitle = element_text(size = 9, color = "gray50"))

print(viz1)
```



Visualization 2: Parallel Coordinates

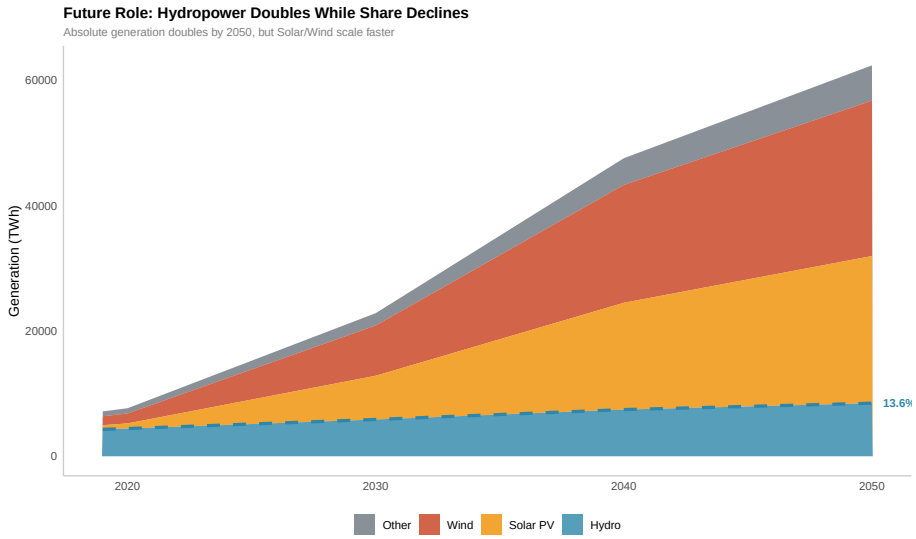
```
# Upstream: med_5, Non-combustion: med_15, Total: med_25
viz2_data <- gleam_clean |>
  filter(!is.na(med_5), !is.na(med_15), !is.na(med_25)) |>
  select(tech_name, tech_type, Upfront = med_5, Ongoing = med_15, Lifetime = med_25) |>
  pivot_longer(cols = c(Upfront, Ongoing, Lifetime), names_to = "dimension", values_to = "va
  mutate(dimension = factor(dimension, levels = c("Upfront", "Ongoing", "Lifetime")))

viz2 <- viz2_data |>
  ggplot(aes(x = dimension, y = value, group = tech_name, color = tech_type)) +
  geom_line(size = 1.5, alpha = 0.8) +
  geom_point(size = 3) +
  geom_text_repel(data = filter(viz2_data, dimension == "Lifetime"),
                  aes(label = tech_name), hjust = -0.1, size = 3, fontface = "bold") +
  annotate("text", x = 1, y = 50, label = "High upfront\nfor hydro",
           color = "#2c7bb6", size = 2.5, hjust = 0) +
  annotate("text", x = 3, y = 800, label = "Fossil fuels\nworse over time",
           color = "#d73027", size = 2.5, hjust = 1) +
  scale_color_manual(values = c("Hydropower" = "#2c7bb6", "Fossil Fuel" = "#d73027", "Nuclea
  scale_y_log10() +
  labs(title = "The Temporal Pattern: Upfront vs. Lifetime",
       subtitle = "Hydropower starts high but stays flat; fossil fuels accumulate",
       x = NULL,
       y = "Emissions (g CO e/kWh)",
       color = NULL) +
  theme_minimal() +
  theme(legend.position = "none",
        panel.grid = element_blank(),
        axis.line = element_line(color = "gray80"),
        plot.title = element_text(face = "bold", size = 12),
        plot.subtitle = element_text(size = 9, color = "gray50"))

print(viz2)
```



Visualization 3: Stacked Area

```
# Load IEA data
nze_data <- read.csv("data/NZE2021_AnnexA.csv")

hydro_data <- nze_data |>
  filter(Product %in% c("Hydro", "Solar PV", "Wind", "Renewables", "Total"),
         Flow == "Electricity generation",
         Year %in% c(2019, 2020, 2030, 2040, 2050))

generation <- hydro_data |>
  filter(Unit == "TWh") |>
  select(Product, Year, Value) |>
  pivot_wider(names_from = Product, values_from = Value) |>
  mutate(
    Other = Renewables - Hydro - `Solar PV` - Wind,
    Hydro_Share = (Hydro / Renewables) * 100
  )

viz3_data <- generation |>
  select(Year, Hydro, `Solar PV`, Wind, Other) |>
  pivot_longer(cols = -Year, names_to = "Source", values_to = "Generation") |>
  mutate(Source = factor(Source, levels = c("Other", "Wind", "Solar PV", "Hydro")))

viz3 <- viz3_data |>
  ggplot(aes(x = Year, y = Generation, fill = Source)) +
  geom_area(alpha = 0.8) +
```

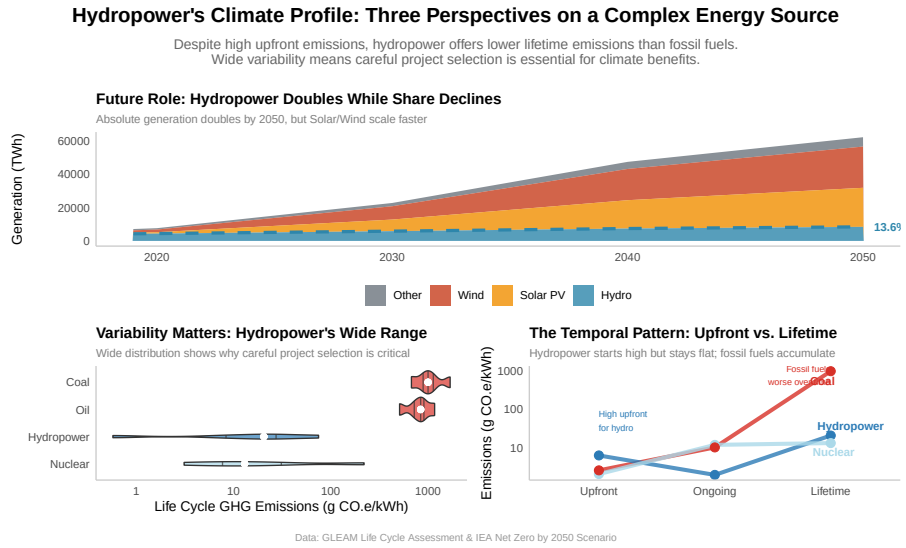
```
  geom_line(data = generation, aes(x = Year, y = Hydro),
            inherit.aes = FALSE, linetype = "dashed", size = 1.2, color = "#2E86AB") +
  geom_text(data = generation |> filter(Year == 2050),
            aes(x = 2050, y = Hydro, label = paste0(round(Hydro_Share, 1), "% of renewables'
            inherit.aes = FALSE, hjust = -0.1, vjust = 0.5, size = 3,
            fontface = "bold", color = "#2E86AB") +
  scale_fill_manual(values = c("Hydro" = "#2E86AB", "Solar PV" = "#F18F01",
                               "Wind" = "#C73E1D", "Other" = "#6C757D")) +
  labs(title = "Future Role: Hydropower Doubles While Share Declines",
       subtitle = "Absolute generation doubles by 2050, but Solar/Wind scale faster",
       x = NULL,
       y = "Generation (TWh)",
       fill = NULL) +
  theme_minimal() +
  theme(legend.position = "bottom",
        panel.grid = element_blank(),
        axis.line = element_line(color = "gray80"),
        plot.title = element_text(face = "bold", size = 12),
        plot.subtitle = element_text(size = 9, color = "gray50"),
        legend.text = element_text(size = 9))

print(viz3)
```



Combined Infographic

```
# Create a unified infographic with connecting narrative
combined <- (viz3 / (viz1 | viz2)) +
```

```
  plot_annotation(
    title = "Hydropower's Climate Profile: Three Perspectives on a Complex Energy Source",
    subtitle = "Despite high upfront emissions, hydropower offers lower lifetime emissions t
    caption = "Data: GLEAM Life Cycle Assessment & IEA Net Zero by 2050 Scenario",
    theme = theme(
      plot.title = element_text(size = 16, face = "bold", hjust = 0.5, margin = margin(b = 1
      plot.subtitle = element_text(size = 11, hjust = 0.5, color = "gray40", margin = margin
      plot.caption = element_text(size = 8, color = "gray60", hjust = 0.5, margin = margin(t
    )
  )

print(combined)
```



Part VI: Answer a Few Last Questions

Question 1: Key Insights

1. Hydropower has high upfront emissions but lower lifetime totals than fossil fuels
2. Wide variability in hydropower emissions means project selection matters
3. Hydropower's share of renewables will decline but absolute generation doubles

Design choices: Used consistent colors, log scales for emissions, combined layout to show all three perspectives together.

Question 2: Challenges

- Excel file had weird headers, had to use skip=2
- Violin plot from quantile data isn't perfect but works
- Combining two different data sources was tricky

Anticipated: Need to polish colors/fonts for final version, might want to add interactivity later.

Question 3: Packages

Used `patchwork` to combine plots and `ggrepel` for labels. Both learned from documentation. Might need `gganimate` or `plotly` later if I want interactivity.

Question 4: Feedback Needed

- Is my scope still to big? Should I narrow it down?
- I spent a lot of time on more exploartoy visuals and wrangling, is there any interesting visualizations you might like to see?
- Is the violin plot approach fine given I only have quantiles?