Analytical Panel — Basic Summary Stats

NHMC Project

October 21, 2025

Contents


```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)

suppressPackageStartupMessages({
  library(tidyverse)
  library(janitor)
  library(glue)
  library(knitr)
  library(kableExtra)
})

# Hardcoded path to the analytical panel
path <- "C:/Repositories/white-bowblis-nhmc/data/clean/analytical_panel.csv"

if (!file.exists(path)) {
  stop(paste0("File not found at: ", path), call. = FALSE)
}

# === Load and clean ===
df <- readr::read_csv(path, show_col_types = FALSE) |>
  clean_names() |>
  # Only convert empty strings to NA for character columns
  mutate(across(where(is.character), ~ na_if(trimws(.), "")))

# Convert year_month to a date if present
if ("year_month" %in% names(df)) {
  df <- df %>%
    mutate(year_month = as.Date(paste0(year_month, "/01"), format = "%Y/%m/%d"))
```

```
}

# (Optional) Drop columns you don't need, e.g.:
# df <- dplyr::select(df, -any_of("time"))
```

1 Quick Overview

```
n_rows   <- nrow(df)
n_ccn    <- dplyr::n_distinct(df$cms_certification_number)
n_months <- dplyr::n_distinct(df$year_month)

glue("Rows: {scales::comma(n_rows)} | Unique CCNs: {scales::comma(n_ccn)} | Unique Months: {n_
```

2 Numeric Variables: Summary Table

```
num_vars <- c(
  "rn_hppd","lpn_hppd","cna_hppd","total_hppd",
  "num_beds","occupancy_rate","pct_medicare","pct_medicaid"
)
num_vars <- intersect(num_vars, names(df))

summ_tbl <- df %>%
  summarise(
    across(
      all_of(num_vars),
      list(
        n    = ~ sum(!is.na(.)),
        mean = ~ mean(., na.rm = TRUE),
        sd   = ~ sd(., na.rm = TRUE),
        p25  = ~ quantile(., 0.25, na.rm = TRUE),
        p50  = ~ quantile(., 0.50, na.rm = TRUE),
        p75  = ~ quantile(., 0.75, na.rm = TRUE)
      ),
      .names = "{.col}**{.fn}"
    )
  ) %>%
  pivot_longer(everything(),
               names_to = c("variable","stat"),
               names_sep = "\\*\\*",
               values_to = "value") %>%
  pivot_wider(names_from = stat, values_from = value) %>%
  mutate(across(where(is.numeric), ~ round(., 3)))

kable(summ_tbl, caption = "Basic summary stats (all rows)") %>%
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position")
```

3 Binary / Flag Variables (Facility-Level "Ever")

```
ever_one <- function(x) {
  x <- as.integer(x)
  if (all(is.na(x))) return(NA_integer_)
  as.integer(any(x == 1, na.rm = TRUE))
}

ever_zero <- function(x) {
  x <- as.integer(x)
  if (all(is.na(x))) return(NA_integer_)
  as.integer(any(x == 0, na.rm = TRUE) & !any(x == 1, na.rm = TRUE))
}

bin_vars <- c(
  "non_profit","government","chain","urban",
  "provider_resides_in_hospital","ccrc_facility","sff_facility"
)
bin_vars <- intersect(bin_vars, names(df))

fac <- df %>%
  group_by(cms_certification_number) %>%
  summarise(
    across(
      all_of(bin_vars),
      list(
        ever1 = ~ ever_one(.),
        ever0 = ~ ever_zero(.),
        miss  = ~ as.integer(all(is.na(.)))
      ),
      .names = "{.col}__{.fn}"
    ),
    .groups = "drop"
  )

count_rows <- purrr::map_dfr(bin_vars, function(v) {
  tibble(
    variable = v,
    n_ever_1 = sum(fac[[paste0(v, "__ever1")]] == 1, na.rm = TRUE),
    n_ever_0 = sum(fac[[paste0(v, "__ever0")]] == 1, na.rm = TRUE),
    n_all_na = sum(fac[[paste0(v, "__miss")]]  == 1, na.rm = TRUE)
  )
}) %>%
  mutate(
    pct_ever_1 = scales::percent(n_ever_1 / n_ccn),
    pct_ever_0 = scales::percent(n_ever_0 / n_ccn),
    pct_all_na = scales::percent(n_all_na / n_ccn)
  )
```

```
kable(count_rows, caption = "Facility-level counts by CCN (ever=1 / ever=0 / all NA)") %>%
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position")
```

4 Highlight: Ever = 1 (by CCN)

```
hl <- count_rows %>%
  select(variable, n_ever_1, pct_ever_1) %>%
  arrange(desc(n_ever_1))

kable(hl, caption = "Facilities with the flag = 1 at least once (by CCN)") %>%
  kable_styling(full_width = FALSE, position = "center", latex_options = "hold_position")
```