```
# Add Customers and percent_customers_out to county_daily_peak.csv
# Output: county_daily_peak_with_percent.csv

# install.packages(c("data.table","janitor","stringr"))
library(data.table)
library(janitor)
library(stringr)

# ---- Paths ----
data_dir <- "~/path/to/folder"   # <-- change me
peak_csv <- file.path(data_dir, "county_daily_peak.csv")
mcc_csv  <- file.path(data_dir, "MCC.csv")
out_csv  <- file.path(data_dir, "county_daily_peak_with_percent.csv")

# ---- Read ----
peak <- fread(peak_csv, na.strings = c("", "NA", "NaN", "null"))
mcc  <- fread(mcc_csv,  na.strings = c("", "NA", "NaN", "null"))

# Normalize header names
setnames(peak, make_clean_names(names(peak)))
setnames(mcc,  make_clean_names(names(mcc)))

# Expecting:
# peak: date, fips_code, county, state, county_daily_peak
# mcc : county_fips, customers
need_peak <- c("date","fips_code","county","state","county_daily_peak")
need_mcc  <- c("county_fips","customers")
miss_p <- setdiff(need_peak, names(peak)); if (length(miss_p)) stop("Missing in
county_daily_peak.csv: ", paste(miss_p, collapse=", "))
miss_m <- setdiff(need_mcc,  names(mcc));  if (length(miss_m)) stop("Missing in MCC.csv:
", paste(miss_m, collapse=", "))

# ---- Types & padding ----
peak[ , fips_code   := str_pad(as.character(fips_code),   width = 5, pad = "0")]
mcc[  , county_fips := str_pad(as.character(county_fips), width = 5, pad = "0")]

peak[, county_daily_peak := as.numeric(gsub(",", "", county_daily_peak))]
mcc[ , customers       := as.numeric(gsub(",", "", customers))]

# ---- Left join (no aggregation; MCC is 1 row per county) ----
setkey(peak, fips_code)
setkey(mcc,  county_fips)
merged <- mcc[peak, on = c(county_fips = "fips_code")]  # keeps all rows from 'peak'
```

```
# ---- Canonicalize FIPS name so downstream code is stable ----
# After the join we have 'county_fips' (from MCC) and not always 'fips_code';
# create a single standard column called 'fips_code'.
if (!"fips_code" %in% names(merged) && "county_fips" %in% names(merged)) {
  setnames(merged, "county_fips", "fips_code")
}

# ---- Compute percent ----
merged[, percent_customers_out := fifelse(!is.na(customers) & customers > 0,
                                          100 * county_daily_peak / customers,
                                          NA_real_)]
merged[, percent_customers_out := round(pmax(0, pmin(100, percent_customers_out)),
4)]

# ---- Arrange columns & write ----
desired <-
c("date","fips_code","county","state","county_daily_peak","customers","percent_customers_
out")
have <- intersect(desired, names(merged))
setcolorder(merged, c(have, setdiff(names(merged), have)))

fwrite(merged, out_csv)
message("Wrote: ", out_csv, " (", nrow(merged), " rows)")

# Optional quick sanity check (now using the canonical fips_code):
# dup_mcc <- merged[, .N, by = .(fips_code)][N > 1, .N]
# if (dup_mcc > 0) message("Note: Some FIPS appear multiple times after join (check
inputs).")
```