

Dates with too many MDJS cases

Dates with too many cases

```
library(here)
library(dplyr)
library(readr)
library(purrr)
library(stringr)
library(data.table)
```

```
# Set directories.
path_to_log_files <- here("output", "log_tables")
path_to_search_table <- here("output", "search_tables")
search_table_name <- "all_counties_0_days_too_many_cases.csv"
```

```
# Read in log files.
log_files <- list.files(path_to_log_files, full.names = T)
log_files <- log_files[str_detect(log_files, "too_many_cases_log_part")]

log_table <-
  map(log_files, function(csv, path) {fread(csv)}) %>%
  bind_rows()
```

```
# Check number of dates which could not be collected.
nr_dates_too_many <- log_table %>% filter(too_many) %>% nrow()

nr_dates_too_many_full <-
  log_table %>% filter(too_many, !mdjs_too_many, !cp_too_many) %>% nrow()

nr_dates_too_many_mdjs <-
  log_table %>% filter(too_many, mdjs_too_many, !cp_too_many) %>% nrow()
```

```
nr_dates_too_many_cp <-
  log_table %>% filter(too_many, !mdjs_too_many, cp_too_many) %>% nrow()

nr_dates_too_many_mdjs_cp <-
  log_table %>% filter(too_many, mdjs_too_many, cp_too_many) %>% nrow()
```

- Number of dates with too many cases: 791
- Number of dates with too many cases but we were still able to collect all the cases: 636
- Number of dates with too many cases and we were not able to collect all the cases (too many MDJS cases): 61
 - Note that this number should really be 60 because in between scrapes some cases were purged. As a result, we were able to obtain every case for October 4, 2019 for Lackawanna county.
- Number of dates with too many cases and we were not able to collect all the cases (too many CP cases): 94
- Number of dates with too many cases and we were not able to collect all the cases (too many CP and MDJS cases): 0

```
# Save the information on which dates had too many cases.
dates_too_many_mdjs_cp <- log_table %>% filter(mdjs_too_many | cp_too_many)
write_csv(
  dates_too_many_mdjs_cp,
  here("documentation", "dates_too_many_mdjs_cp.csv")
)

# Save the information on which dates had too many MDJS cases for final scrape.
dates_too_many_mdjs <-
  log_table %>%
  filter(mdjs_too_many) %>%
  select(-matches("too_many"))

write_csv(
  dates_too_many_mdjs,
  here(path_to_search_table, "all_counties_0_days_too_many_mdjs.csv")
)

# Save the concatenated log file.
write_csv(
  log_table, here(path_to_log_files, "all_counties_0_days_too_many_cases_log.csv")
)
```