Final project

PPHA 30110 | Coding Lab for Public Policy

Jay Ballesteros

November 16, 2025

Introduction

In this final project for the coding lab course, I will analyze the *UCPD dataset* (scrapped by Chicago Maroon) to explore patterns in campus crime reports. The dataset includes various types of incidents reported on campus, along with details such as date, time, location, and description of the incidents. For simplicity, the graph below illustrates the trend of the top 5 incidents reported on campus during 2025 (the last 10 months).

The results shows that specifically `theft` and its subcategory `information / theft` lead the number of reports during this period, with a noticeable peaks in June and October. Other incidents like `information`, `medical call`, and `found property` also show high incidence but with less noticeable trends.

As a side note (technical-no so technical), it is worth to mention that the variable selected for data is `reported_date`. Whereas there is also a variable called `occurred` which could be more relevant to analyze when the crimes actually happened, has limitations in its format, since all dates and minutes have heterogeneous formats. Therefore, I decided to focus on `reported_date` for this analysis.

Graph

```
library(tidyverse)
library(knitr)
library(scales)
df <- read_csv("_data/ucpd_data.csv")
df <- df %>%
  filter(reported_date >= as.Date("2025-01-01") & reported_date <= as.Date("2025-10-31"))

df_month_incident <- df %>%
  mutate(
    reported_month = month(reported_date, label = TRUE, abbr = TRUE),
    incident_group = fct_lump_n(incident, n = 5, other_level = "Other")) %>%
  filter(
    incident_group != "Other") %>%  #here i filtered to not show others because the graph gets a little
  count(
    reported_month, incident_group, name = "count")

ggplot(df_month_incident,
       aes(x = reported_month,
           y = count,
           group = incident_group,
           color = incident_group)) +
  geom_line(linewidth = 1) +
  geom_point(size = 1) +
  labs(
    title = "Monthly number of reports by top 5 incident types during 2025 (Jan-Oct)",
    x = "Reported month",
    y = "Number of reports",
    color = "Incident"
  ) +
  theme_minimal()
```
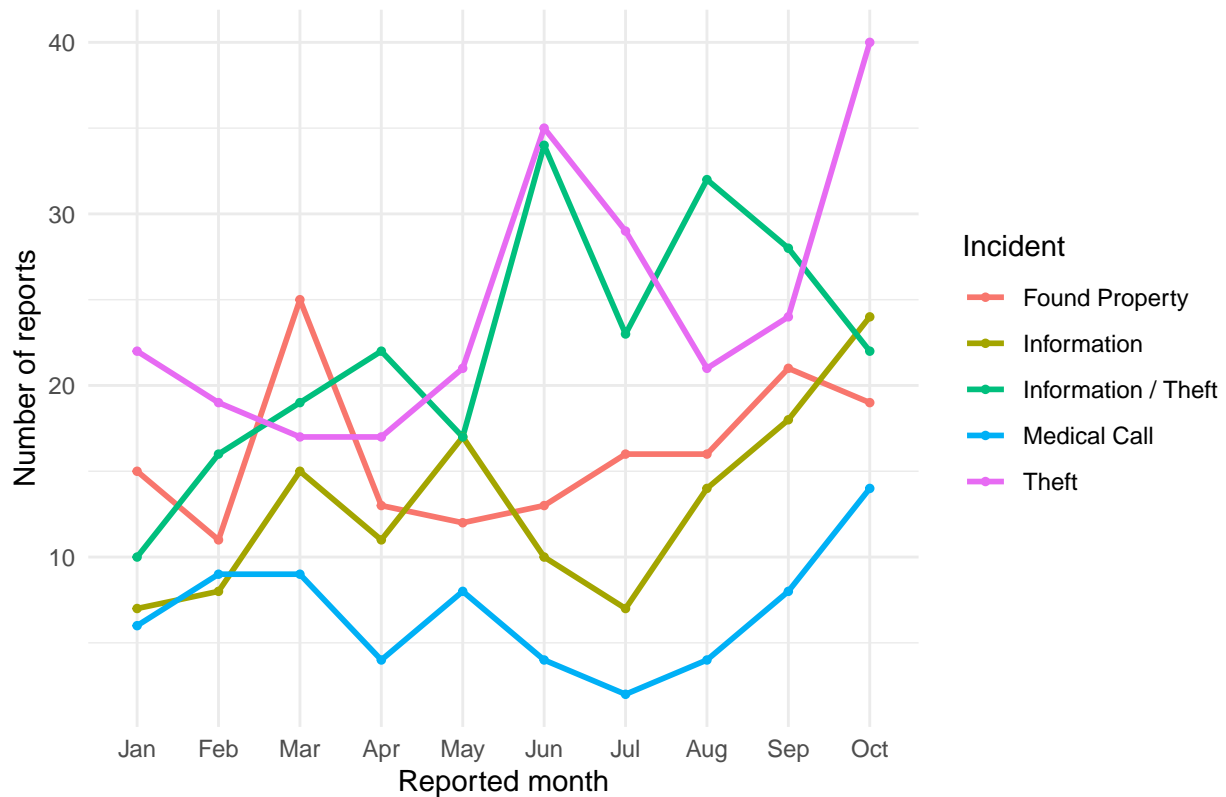
Monthly number of reports by top 5 incident types during 2025 (Jan–Oct)

Table

To better show proportions of main incidents by month, in the following table I expand to showcase the top 10 incidents reported on campus during the same period and grouping the remaining ones as `Other`.

```r
df_month_incident <- df %>%
  mutate(
    reported_month = month(reported_date, label = TRUE, abbr = TRUE),
    incident_group = fct_lump_n(incident, n = 10, other_level = "Other incidents")) %>%
  count(
    reported_month, incident_group, name = "count")

summary_table <- df_month_incident %>%
  group_by(reported_month) %>%
  mutate(
    month_total = sum(count),
    month_proportion = count / month_total
  ) %>%
  arrange(reported_month, desc(count))

summary_table %>%
  knitr::kable()
```

reported_month	incident_group	count	month_total	month_proportion
Jan	Other incidents	66	141	0.4680851

reported_month	incident_group	count	month_total	month_proportion
Jan	Theft	22	141	0.1560284
Jan	Found Property	15	141	0.1063830
Jan	Information / Theft	10	141	0.0709220
Jan	Information	7	141	0.0496454
Jan	Medical Call	6	141	0.0425532
Jan	Traffic Crash / Property Damage / Hit and Run	6	141	0.0425532
Jan	Medical Transport	4	141	0.0283688
Jan	Battery	2	141	0.0141844
Jan	Found Narcotic	2	141	0.0141844
Jan	Liquor Law Violation	1	141	0.0070922
Feb	Other incidents	59	140	0.4214286
Feb	Theft	19	140	0.1357143
Feb	Information / Theft	16	140	0.1142857
Feb	Found Property	11	140	0.0785714
Feb	Medical Call	9	140	0.0642857
Feb	Battery	8	140	0.0571429
Feb	Information	8	140	0.0571429
Feb	Liquor Law Violation	3	140	0.0214286
Feb	Medical Transport	3	140	0.0214286
Feb	Found Narcotic	2	140	0.0142857
Feb	Traffic Crash / Property Damage / Hit and Run	2	140	0.0142857
Mar	Other incidents	63	177	0.3559322
Mar	Found Property	25	177	0.1412429
Mar	Information / Theft	19	177	0.1073446
Mar	Theft	17	177	0.0960452
Mar	Information	15	177	0.0847458
Mar	Battery	11	177	0.0621469
Mar	Medical Call	9	177	0.0508475
Mar	Medical Transport	7	177	0.0395480
Mar	Found Narcotic	6	177	0.0338983
Mar	Liquor Law Violation	3	177	0.0169492
Mar	Traffic Crash / Property Damage / Hit and Run	2	177	0.0112994
Apr	Other incidents	78	171	0.4561404
Apr	Information / Theft	22	171	0.1286550
Apr	Theft	17	171	0.0994152
Apr	Found Property	13	171	0.0760234
Apr	Information	11	171	0.0643275
Apr	Medical Transport	8	171	0.0467836
Apr	Battery	5	171	0.0292398
Apr	Found Narcotic	5	171	0.0292398
Apr	Traffic Crash / Property Damage / Hit and Run	5	171	0.0292398
Apr	Medical Call	4	171	0.0233918
Apr	Liquor Law Violation	3	171	0.0175439
May	Other incidents	81	189	0.4285714
May	Theft	21	189	0.1111111
May	Information	17	189	0.0899471
May	Information / Theft	17	189	0.0899471
May	Found Property	12	189	0.0634921

reported_month	incident_group	count	month_total	month_proportion
May	Medical Call	8	189	0.0423280
May	Medical Transport	8	189	0.0423280
May	Traffic Crash / Property Damage / Hit and Run	8	189	0.0423280
May	Battery	7	189	0.0370370
May	Found Narcotic	7	189	0.0370370
May	Liquor Law Violation	3	189	0.0158730
Jun	Other incidents	89	206	0.4320388
Jun	Theft	35	206	0.1699029
Jun	Information / Theft	34	206	0.1650485
Jun	Found Property	13	206	0.0631068
Jun	Information	10	206	0.0485437
Jun	Traffic Crash / Property Damage / Hit and Run	7	206	0.0339806
Jun	Medical Transport	6	206	0.0291262
Jun	Battery	4	206	0.0194175
Jun	Found Narcotic	4	206	0.0194175
Jun	Medical Call	4	206	0.0194175
Jul	Other incidents	67	163	0.4110429
Jul	Theft	29	163	0.1779141
Jul	Information / Theft	23	163	0.1411043
Jul	Found Property	16	163	0.0981595
Jul	Found Narcotic	7	163	0.0429448
Jul	Information	7	163	0.0429448
Jul	Traffic Crash / Property Damage / Hit and Run	6	163	0.0368098
Jul	Battery	5	163	0.0306748
Jul	Medical Call	2	163	0.0122699
Jul	Medical Transport	1	163	0.0061350
Aug	Other incidents	68	172	0.3953488
Aug	Information / Theft	32	172	0.1860465
Aug	Theft	21	172	0.1220930
Aug	Found Property	16	172	0.0930233
Aug	Information	14	172	0.0813953
Aug	Found Narcotic	5	172	0.0290698
Aug	Traffic Crash / Property Damage / Hit and Run	5	172	0.0290698
Aug	Battery	4	172	0.0232558
Aug	Medical Call	4	172	0.0232558
Aug	Medical Transport	3	172	0.0174419
Sep	Other incidents	67	194	0.3453608
Sep	Information / Theft	28	194	0.1443299
Sep	Theft	24	194	0.1237113
Sep	Found Property	21	194	0.1082474
Sep	Information	18	194	0.0927835
Sep	Liquor Law Violation	8	194	0.0412371
Sep	Medical Call	8	194	0.0412371
Sep	Battery	7	194	0.0360825
Sep	Medical Transport	7	194	0.0360825
Sep	Found Narcotic	3	194	0.0154639
Sep	Traffic Crash / Property Damage / Hit and Run	3	194	0.0154639

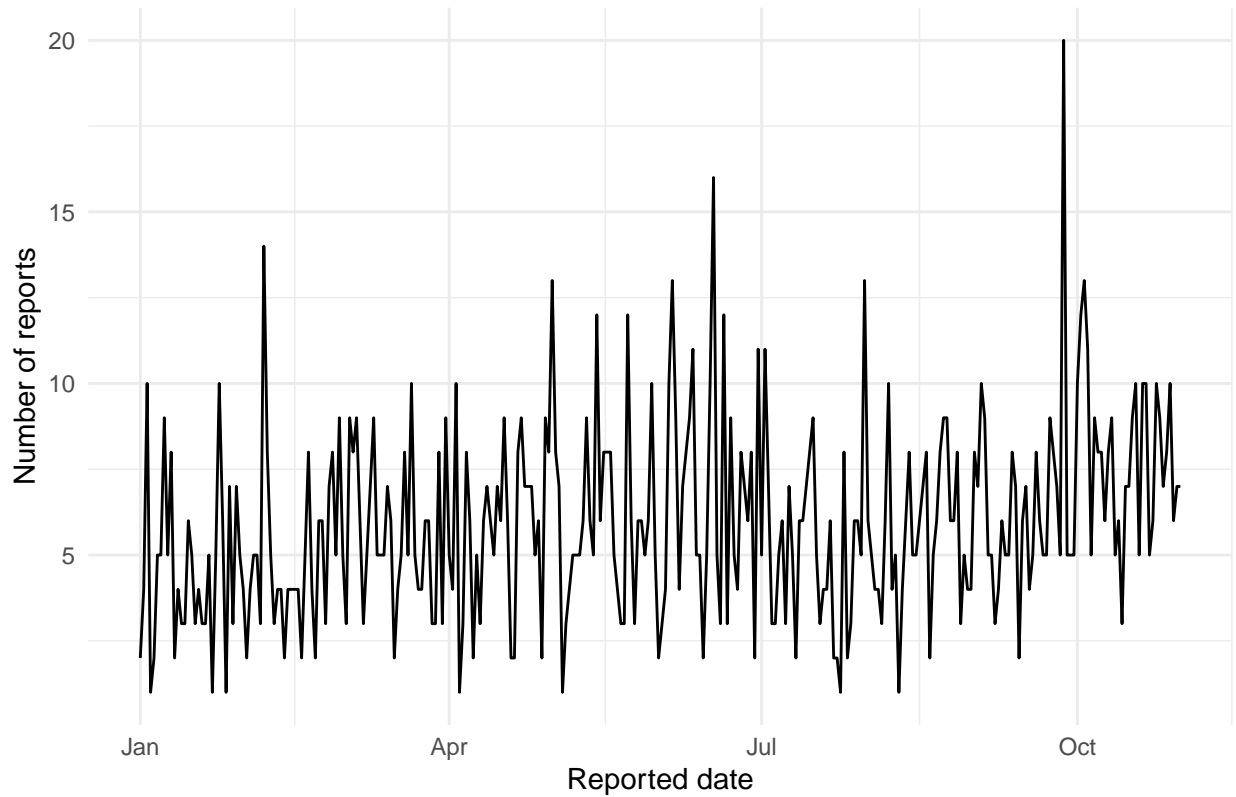| reported_month	incident_group	count	month_total	month_proportion
Oct	Other incidents	88	246	0.3577236
Oct	Theft	40	246	0.1626016
Oct	Information	24	246	0.0975610
Oct	Information / Theft	22	246	0.0894309
Oct	Found Property	19	246	0.0772358
Oct	Liquor Law Violation	16	246	0.0650407
Oct	Medical Call	14	246	0.0569106
Oct	Battery	13	246	0.0528455
Oct	Medical Transport	6	246	0.0243902
Oct	Traffic Crash / Property Damage / Hit and Run	3	246	0.0121951
Oct	Found Narcotic	1	246	0.0040650

Appendix

A funny thing that I noticed when I was doing the exploratory analysis, was that the day with the most reports was September 27, 2025 — this was the first Saturday after back-to-school. The main driver was the incidence of liquor law violations.

```
df_daily_reports <- df %>%
  count(reported_date, name = "count")

ggplot(df_daily_reports,
       aes(x = reported_date,
           y = count)) +
  geom_line() +
  labs(
    title = "Daily number of reports during 2025 (Jan-Oct)",
    x = "Reported date",
    y = "Number of reports"
  ) +
  theme_minimal()
```

Daily number of reports during 2025 (Jan–Oct)


```r
df_sep27 <- df %>%
  filter(reported_date == as.Date("2025-09-27")) %>%
  count(incident, name = "count") %>%
  mutate(
    total_day = sum(count),
    proportion_day = count / total_day
  ) %>%
  arrange(desc(count))

df_sep27 %>%
  knitr::kable()
```

incident	count	total_day	proportion_day
Liquor Law Violation	5	20	0.25
Found Property	3	20	0.15
Medical Call	3	20	0.15
Theft	2	20	0.10
Criminal Damage to Vehicle	1	20	0.05
Information / Damage to Property	1	20	0.05
Information / Domestic Battery	1	20	0.05
Information / Trespass to Property	1	20	0.05
Sexual Assault	1	20	0.05
Theft / Damage to Property	1	20	0.05
Trespass to Property	1	20	0.05