Problemset5

Sharon Allman, Diego Mamanche Castellanos & Ke-Li Chiu

11/04/2020

Abstract

The objective of this analysis was to examine if there was a relationship between [...]

Introduction (Sharon)

Canadian context

https://www.sciencedaily.com/releases/2016/08/160831102834.htm

High-alcohol usage crimes (reference study)

https://www-degruyter-com.myaccess.library.utoronto.ca/view/journals/bejeap/14/3/article-p791.xml

Research Question & Hypotheses

This study intend to confirm [...]

Dataset Description and Data Cleaning

The dataset is the replication data of The Minimum Legal Drinking Age and Crime by Christopher Carpenter and Carlos Dobkin (2015). The dataset contains 1461 observations and 144 variables. The observation is arrestee's age in days relative to 21 on the day being arrested ranging from 17 to 24. For example, 0 indicates that the arrestee is a month away from turn 21, 60 indicates that the arrestee is 21 and two months old and 0 means the arrestee is arrested on their 21st birthday. The crimes data retrieved from California's Monthly Arrest and Citation Register dated from 1979 to 2006, containing major crime types categorized by FBI: violent crime, alcohol-related offenses, property crime, illegal drug possession or sale, and all other offenses. The dataset contains the counts of arrests as well the arrest rates per 10,000 person-years for each crime for each observation.

Method

To validate the impact of the Minimum Legal Drinking Age (MLDA) over crimes, we conducted Regression Discontinuity Design (RDD). According to the researchers in our reference study, after applying their research through several surveys to inmates and other techniques, they found the fifteen most affected crimes by alcohol consumption. For this study, we used the crimes classified as High-alcohol usage. Those are Assault, Rape, Arson, and Murder, based on the role alcohol plays when an offender commits a given crime. Then,

Summary Statistics

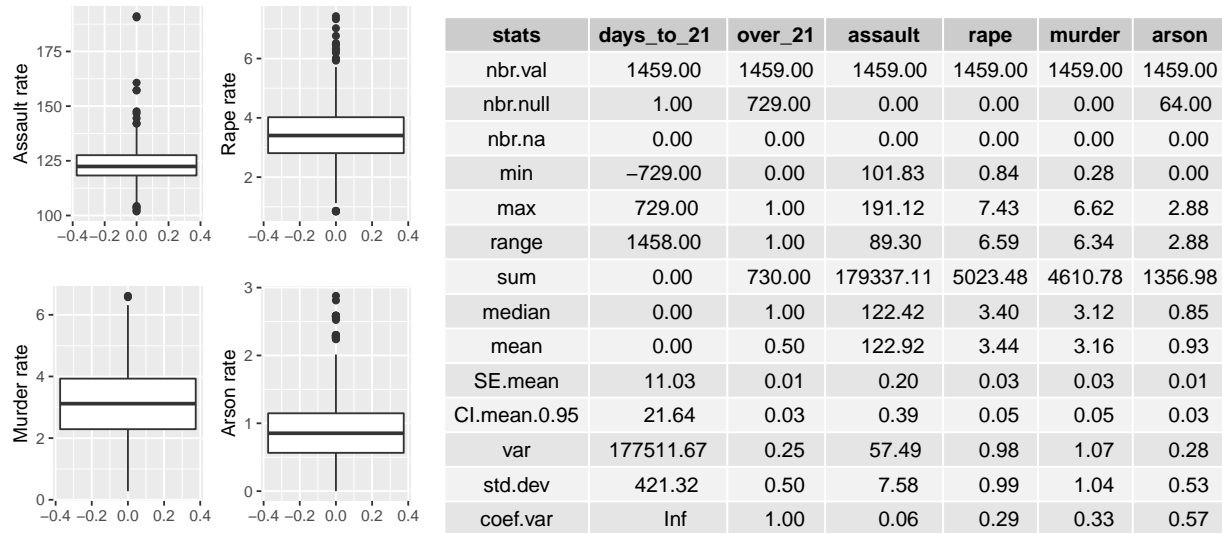The distribution of the High-alcohol usage crimes is shown in Figure 1.

stats	days_to_21	over_21	assault	rape	murder	arson
nbr.val	1459.00	1459.00	1459.00	1459.00	1459.00	1459.00
nbr.null	1.00	729.00	0.00	0.00	0.00	64.00
nbr.na	0.00	0.00	0.00	0.00	0.00	0.00
min	−729.00	0.00	101.83	0.84	0.28	0.00
max	729.00	1.00	191.12	7.43	6.62	2.88
range	1458.00	1.00	89.30	6.59	6.34	2.88
sum	0.00	730.00	179337.11	5023.48	4610.78	1356.98
median	0.00	1.00	122.42	3.40	3.12	0.85
mean	0.00	0.50	122.92	3.44	3.16	0.93
SE.mean	11.03	0.01	0.20	0.03	0.03	0.01
CI.mean.0.95	21.64	0.03	0.39	0.05	0.05	0.03
var	177511.67	0.25	57.49	0.98	1.07	0.28
std.dev	421.32	0.50	7.58	0.99	1.04	0.53
coef.var	Inf	1.00	0.06	0.29	0.33	0.57

Figure 1: Descriptive statistics of the data

Ethical Issues

Weaknesses and Limitations

https://journals-scholarsportal-info.myaccess.library.utoronto.ca/details/01953613/v36i0002/103_acdacinyc.xml

Alcohol intake we don't know

As these authors acknowledge in their literature review, a drawback of the age-related research is that it relies on arrest data in order to ascertain age, and it is possible that the effect of alcohol on arrests is not due to crime commission, but to an increased chance of getting caught.

The primary one is that although our approach recovers a causal estimate of the reduced-form effect of the MLDA on arrests, we cannot provide direct evidence on the underlying mechanisms or clean estimates of the persistence of the effects. Though previous research used RDD methods and various survey data sets to show that alcohol consumption increases sharply at the MLDA (Carpenter & Dobkin, 2009; SAMHSA/OAS, 2009; Crost & Guerrero, 2012) and this is likely the primary cause of the increase, other factors might also contribute to the increase in arrest rates we document in this paper.[11]

A second possibility is that people over the drinking age may be more likely to drink in public places such as bars, where crimes may be more likely to result in an arrest (i.e., we may observe changes in arrest behavior that simply reflect changes in venue).

Because violent crimes – particularly murder, manslaughter, and rape – are relatively uncommon, we are limited somewhat by statistical precision when we disaggregate by specific violent crime type.[30]
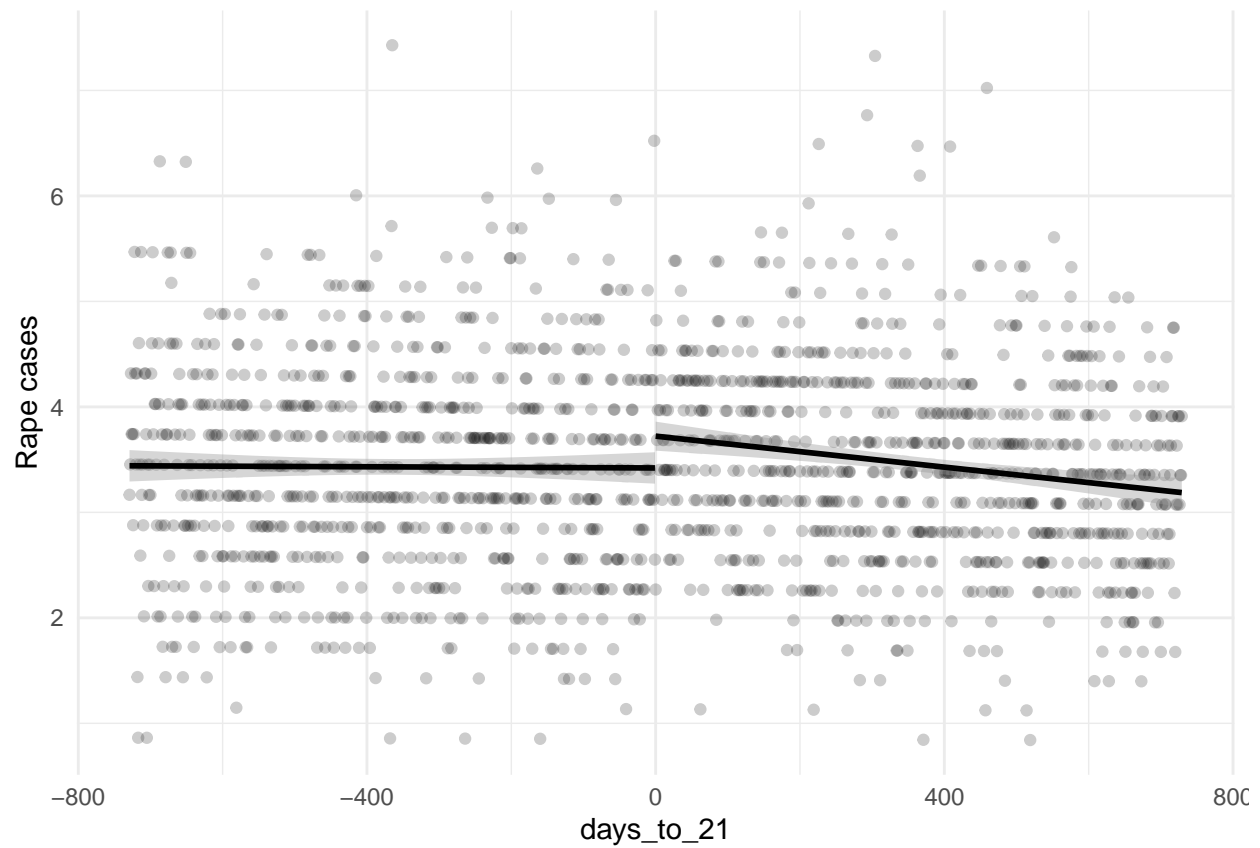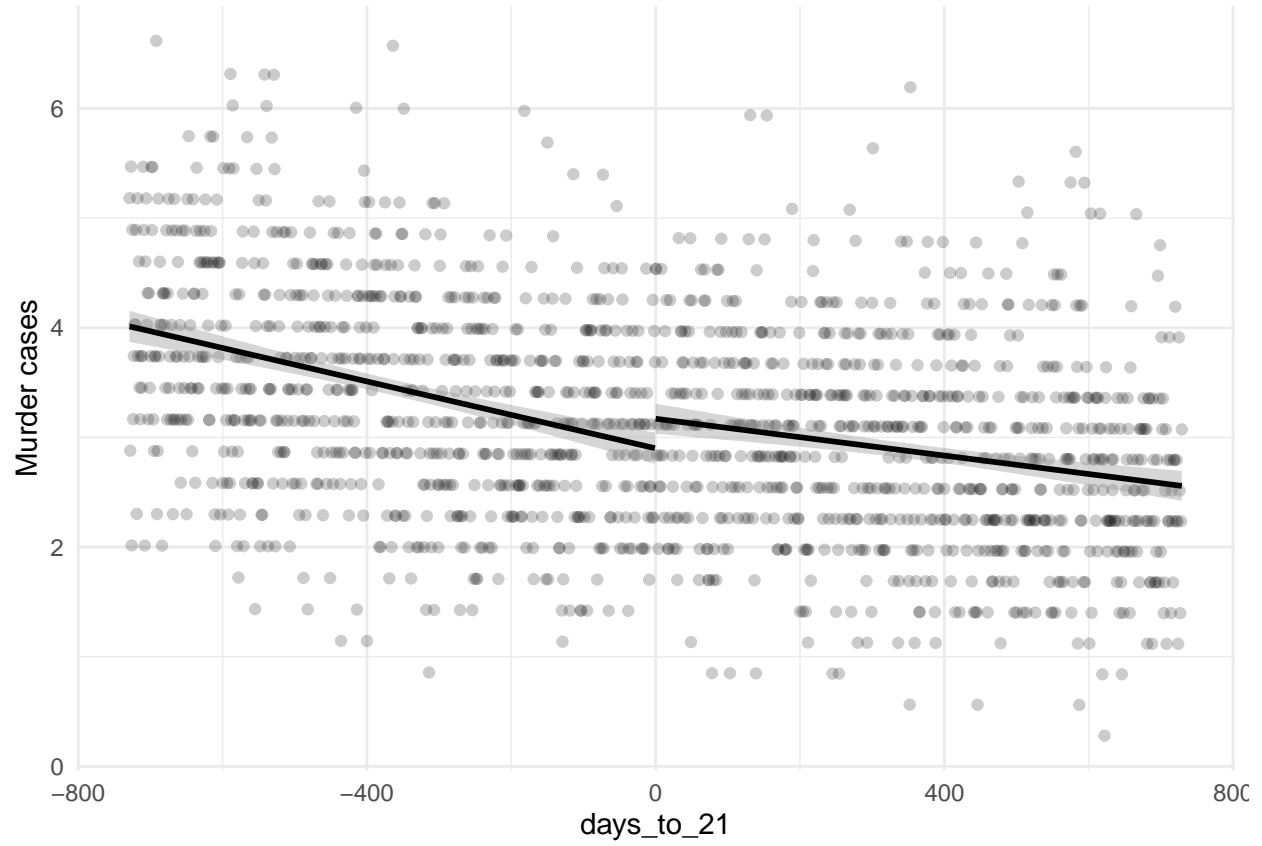
Lack of instrumental variable

Analysis

Assault
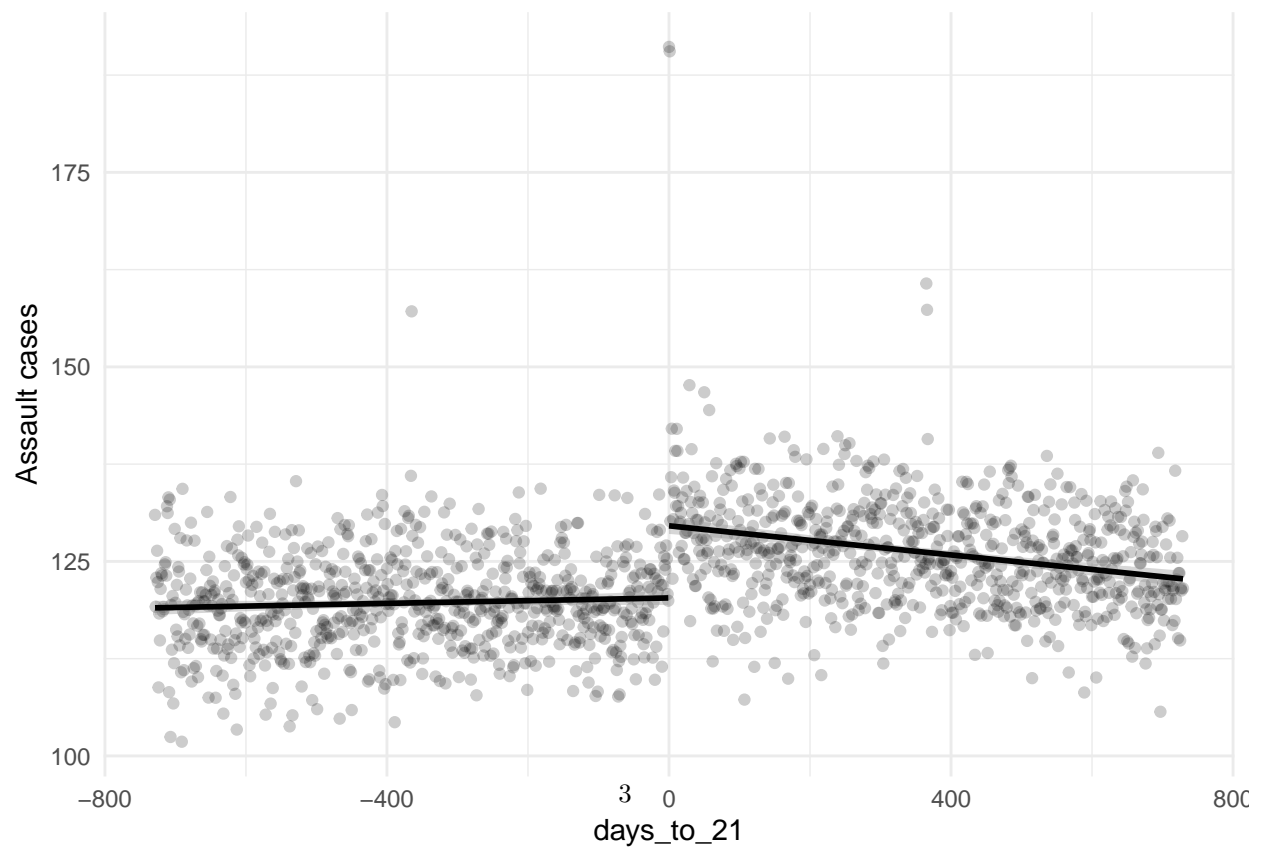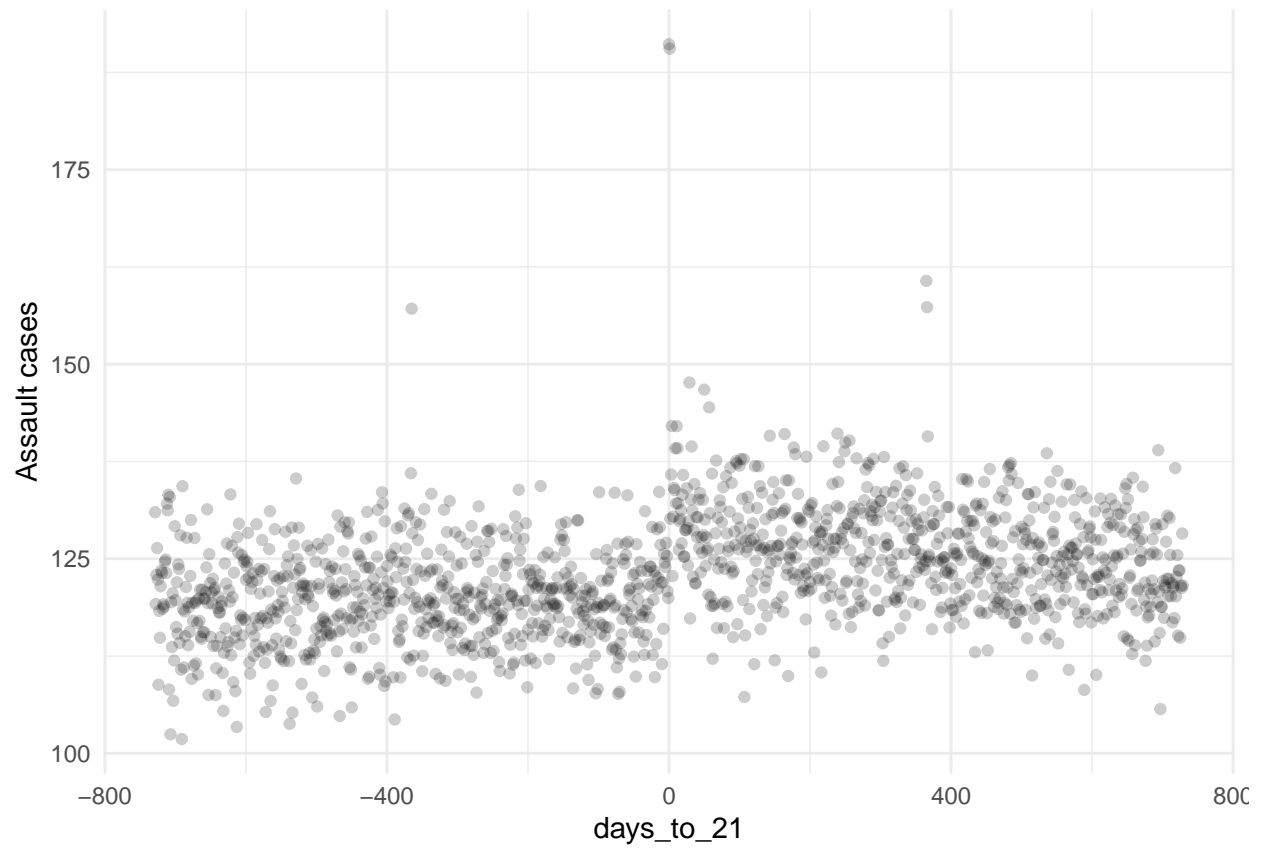
term	estimate	std.error	statistic	p.value
(Intercept)	118	0.399	297	0
days_to_21	-0.00381	0.000846	-4.51	7.13e-06
over_21	9.28	0.713	13	9.9e-37

Rape

term	estimate	std.error	statistic	p.value
(Intercept)	3.29	0.0579	56.8	0
days_to_21	-0.000382	0.000123	-3.11	0.00194
over_21	0.302	0.104	2.92	0.00358

Murder

term	estimate	std.error	statistic	p.value
(Intercept)	3.03	0.0564	53.7	0
days_to_21	-0.00118	0.00012	-9.88	2.56e-22
over_21	0.269	0.101	2.67	0.00763

Arson
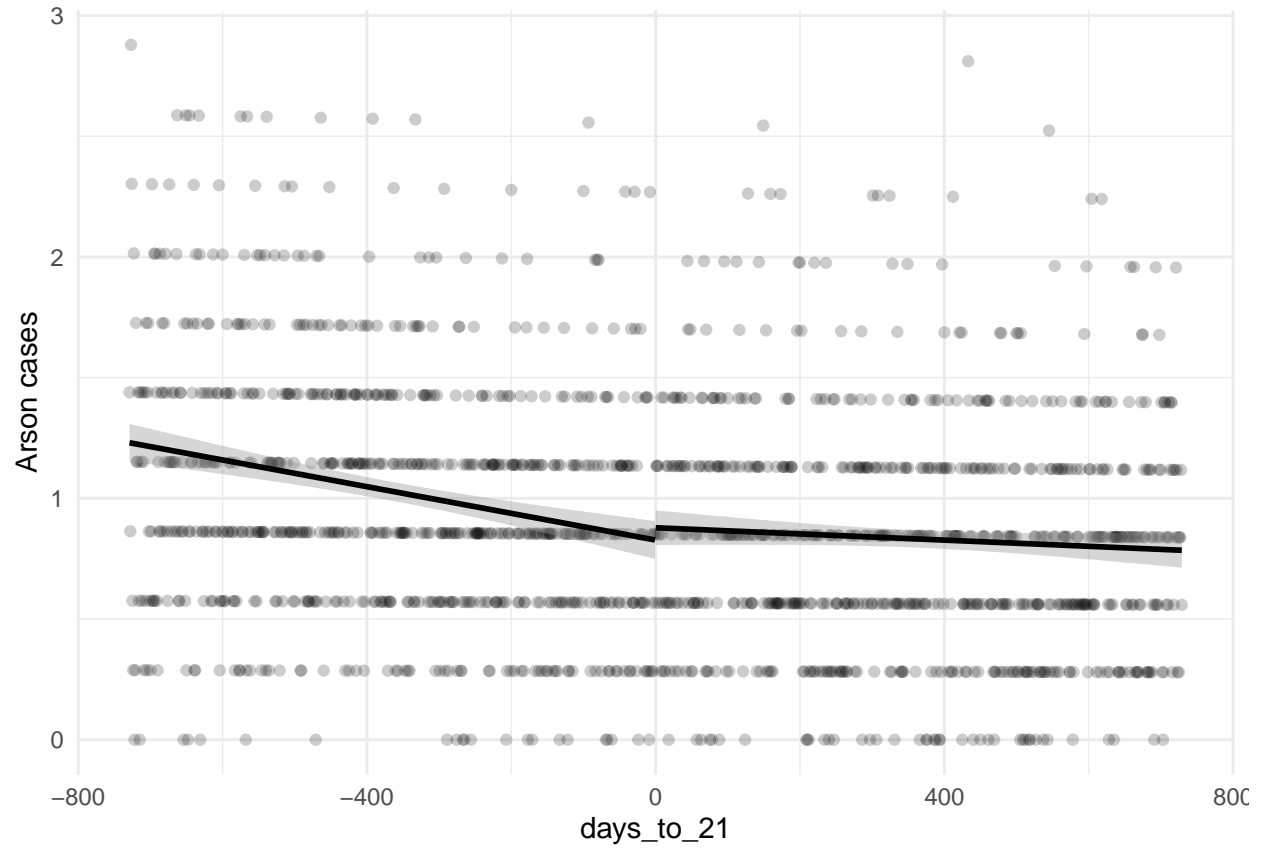
term	estimate	std.error	statistic	p.value
(Intercept)	0.905	0.0304	29.8	8.29e-153
days_to_21	-0.00034	6.44e-05	-5.28	1.49e-07
over_21	0.0506	0.0543	0.933	0.351

Appendix A

```r
knitr::opts_chunk$set(echo = FALSE, include = FALSE)
# import libraries
library(broom)
library(tidyverse)
library(gridExtra)
library(pastecs)
library(readxl)
library(dplyr)
library(tidyr)
library(ggplot2)
library(reshape2)
library(stringr)
library(knitr)
library(huxtable)
library(jtools)

setwd("~/Experimental Design for Data Science/ProblemSet5")
drinking <- read.csv("P01 Age Profile of Arrest Rates 1979-2006.csv")
drinking <- janitor::clean_names(drinking)

head(drinking)
top4_crimes_19_to_23 <- filter(drinking, days_to_21 > -730 &
    days_to_21 < 730)
top4_crimes_19_to_23 <- select(top4_crimes_19_to_23, days_to_21,
    assault_r, rape_r, murder_r, arson_r)

top4_crimes_19_to_23 <- top4_crimes_19_to_23 %>% mutate(over_21 = if_else(days_to_21 <
    0, 0, 1))


# Compute descriptive statistics - boxplots
p1 <- ggplot(top4_crimes_19_to_23) + aes(y = assault_r) + geom_boxplot() +
    labs(x = "", y = "Assault rate")
p2 <- ggplot(top4_crimes_19_to_23) + aes(y = rape_r) + geom_boxplot() +
    labs(x = "", y = "Rape rate")
p3 <- ggplot(top4_crimes_19_to_23) + aes(y = murder_r) + geom_boxplot() +
    labs(x = "", y = "Murder rate")
p4 <- ggplot(top4_crimes_19_to_23) + aes(y = arson_r) + geom_boxplot() +
    labs(x = "", y = "Arson rate")
grid1 <- grid.arrange(p1, p2, p3, p4, ncol = 2, nrow = 2)
# Select only top4
only_top4 <- top4_crimes_19_to_23
# Compute descriptive statistics - table
tt1 <- ttheme_default()

stats_table <- stat.desc(top4_crimes_19_to_23)
stats_table <- round(stats_table, 2)
stats_table <- mutate(stats_table, stats = row.names(stats_table))
stats_table <- select(stats_table, stats, days_to_21, over_21,
    assault_r, rape_r, murder_r, arson_r)
colnames(stats_table) <- c("stats", "days_to_21 ", "over_21",
```

```r
    "assault", "rape", "murder", "arson")
grid2 <- grid.arrange(tableGrob(stats_table, theme = tt1, rows = NULL),
    ncol = 1, nrow = 1)

# Join the boxplots and the table
grid.arrange(arrangeGrob(grid1, ncol = 1, nrow = 1), arrangeGrob(grid2,
    ncol = 1, nrow = 1), heights = c(15, 1), widths = c(1, 2),
    bottom = "Figure 1: Descriptive statistics of the data")

top4_crimes_19_to_23 %>% ggplot(aes(x = days_to_21, y = assault_r)) +
    geom_point(alpha = 0.2) + theme_minimal() + labs(x = "days_to_21",
    y = "Assault cases")


# Plot assault cases
top4_crimes_19_to_23 %>% ggplot(aes(x = days_to_21, y = assault_r)) +
    geom_point(alpha = 0.2) + geom_smooth(data = top4_crimes_19_to_23 %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = top4_crimes_19_to_23 %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "Assault cases")


lm(assault_r ~ days_to_21 + over_21, data = top4_crimes_19_to_23) %>%
    tidy()

# Plot assault cases
top4_crimes_19_to_23 %>% ggplot(aes(x = days_to_21, y = rape_r)) +
    geom_point(alpha = 0.2) + geom_smooth(data = top4_crimes_19_to_23 %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = top4_crimes_19_to_23 %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "Rape cases")


lm(rape_r ~ days_to_21 + over_21, data = top4_crimes_19_to_23) %>%
    tidy()

# Plot assault cases
top4_crimes_19_to_23 %>% ggplot(aes(x = days_to_21, y = murder_r)) +
    geom_point(alpha = 0.2) + geom_smooth(data = top4_crimes_19_to_23 %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = top4_crimes_19_to_23 %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "Murder cases")


lm(murder_r ~ days_to_21 + over_21, data = top4_crimes_19_to_23) %>%
    tidy()

# Plot assault cases
top4_crimes_19_to_23 %>% ggplot(aes(x = days_to_21, y = arson_r)) +
    geom_point(alpha = 0.2) + geom_smooth(data = top4_crimes_19_to_23 %>%
```

```
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = top4_crimes_19_to_23 %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "Arson cases")


lm(arson_r ~ days_to_21 + over_21, data = top4_crimes_19_to_23) %>%
    tidy()
```

References