

Lab 4: A muRdeR mystery

Elijah Russell

10/02/25

Setup

In this lab, you'll get some experience making different types of plots by using different `geom_` functions in `ggplot`.

Load libraries and the data

I've included some code below to load the data we'll use. This chunk also includes the chunk option `echo=FALSE` which hides it in your final submission. *Don't edit the code in this chunk.*

Solutions

Add code chunks (CMD + OPTION + I on Mac; see more `Rmd` keyboard shortcuts here) to practice and modify any of the functions or lines of code introduced in the tutorial.

Remember that at any time, you can Preview or Knit your `Rmd` file to check that everything will work as intended before you submit your assignment to be graded.

```
# view(crime_scene_report)
# view(facebook_event_checkin)
# view(get_fit_now_member)
# view(interview)
# view(drivers_license)
# view(get_fit_now_checkin)
# view(income)
# view(person)
```

1. Use the `ls()` command to list out all of the datasets loaded in your workspace. How many datasets are there?

There are 8 datasets

```
ls()

## [1] "crime_scene_report"      "drivers_license"          "facebook_event_checkin"
## [4] "get_fit_now_checkin"    "get_fit_now_member"      "income"
## [7] "interview"              "person"
```

2. Which dataset is the largest (here, most rows) and what are its dimensions?

The largest dataset is "facebook_event_checkin". Its dimensions are 20011x4.

```
nrow(crime_scene_report)
```

```
## [1] 1228
```

```
nrow(facebook_event_checkin)
```

```
## [1] 20011
```

```
nrow(get_fit_now_member)
```

```
## [1] 184
```

```
nrow(interview)
```

```
## [1] 4991
```

```
nrow(drivers_license)
```

```
## [1] 10007
```

```
nrow(get_fit_now_checkin)
```

```
## [1] 2703
```

```
nrow(income)
```

```
## [1] 7514
```

```
nrow(person)
```

```
## [1] 10011
```

```
dim(facebook_event_checkin)
```

```
## [1] 20011      4
```

3. What types of crimes are recorded in the `crime_scene_report` table? How many records of each type are there?

The types of crimes are arson, assault, blackmail, bribery, fraud, murder, robbery, smuggling, and theft. There are 148, 145, 130, 135, 130, 148, 134, 117, 141 of each type, respectively.

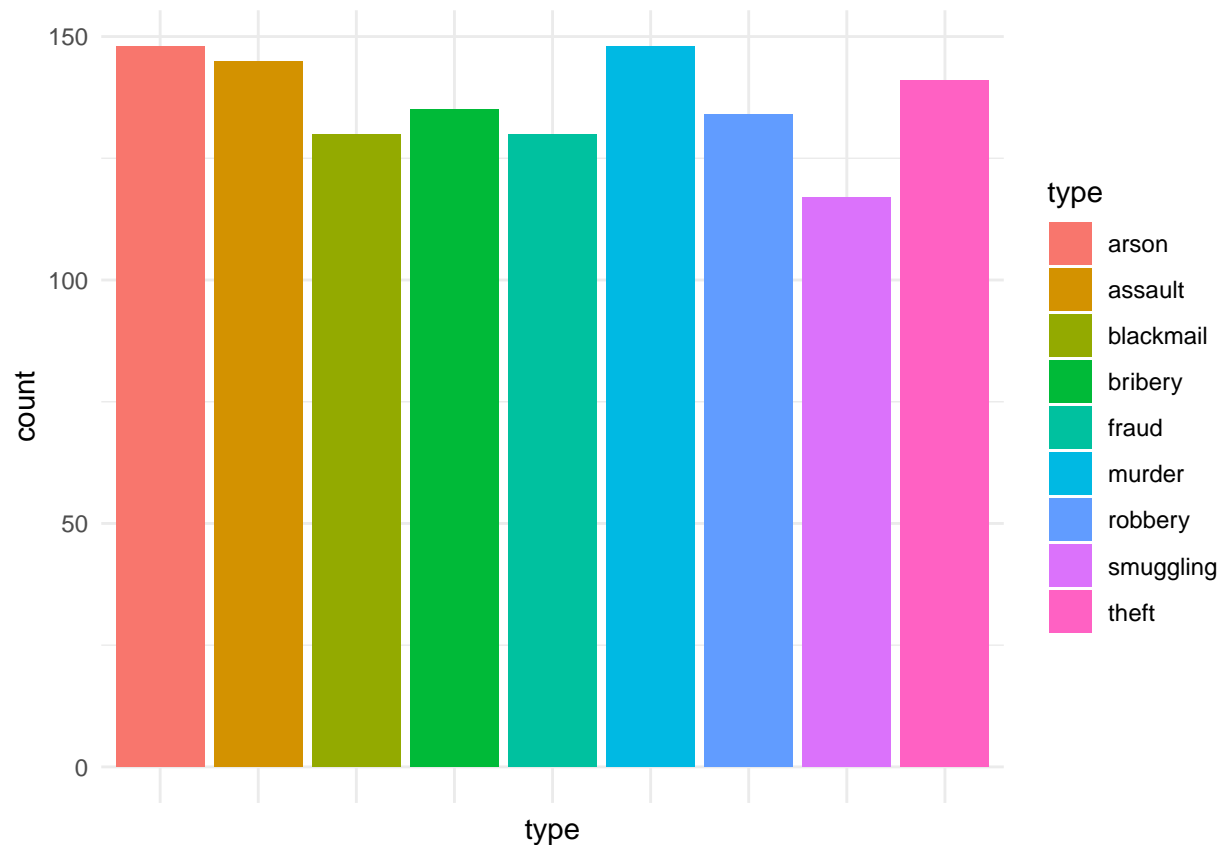
```
table(crime_scene_report$type)
```

```
##
##      arson      assault blackmail      bribery      fraud      murder      robbery smuggling
##       148         145         130         135         130         148         134         117
##      theft
##       141
```

4. Make a bar plot of crime counts by type.

```
library(tidyverse)
library(ggplot2)
library(tidyr)

ggplot(crime_scene_report, aes(x=type, fill=type)) +
  geom_bar() +
  theme_minimal() +
  theme(axis.text.x = element_blank())
```



5. Using the details about the crime given in the very first paragraph at the top of this file, can you change the code above (the line with the logic symbols) to get your first clue?

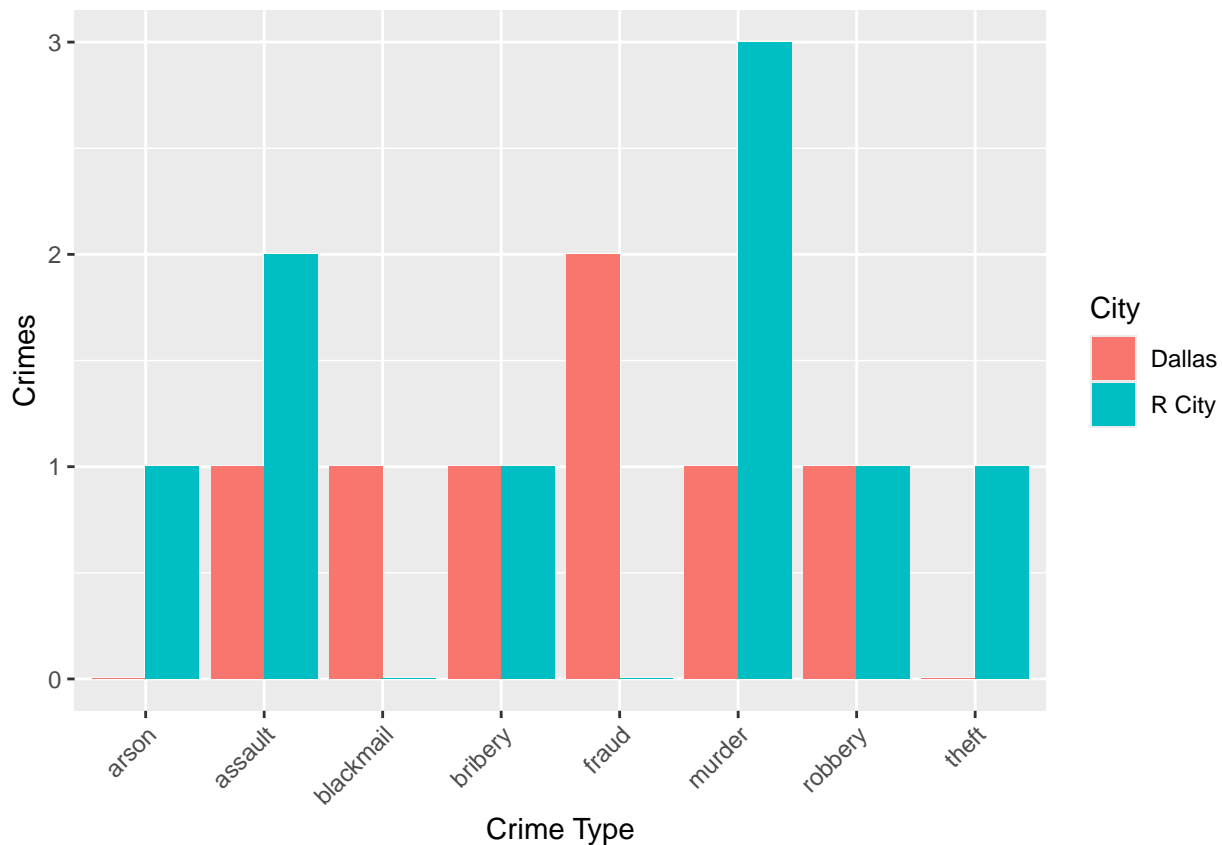
```
first_clue <- crime_scene_report %>%
  filter(type == "murder" & date == "20180115" & city == "R City")

# View(first_clue)
```

6. Make a bar plot that compares crime counts by type in R City to another city of your choosing from the list below?

(Ungraded question): Why have I chosen these cities? idk

```
crime_scene_report %>%
  filter(city %in% c("R City", "Dallas")) %>%
  count(type, city) %>%
  complete(type, city, fill = list(n = 0)) %>%
  ggplot(aes(x=type, y=n, fill=city)) +
  geom_bar(stat="identity", position="dodge") +
  labs(x="Crime Type", y="Crimes", fill="City") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



7. What are the names of the two witnesses?

Morty Schapiro and Annabel Miller

```
witnesses <- person %>%
  filter(address_street_name == "Northwestern Dr" | address_street_name == "Franklin Ave" & first_name == "Morty")
  arrange(address_number) %>%
  select(name, address_street_name, address_number, id)

# View(witnesses)
```

8. Print out the interview transcripts for each of the witnesses.

```
witness_interviews <- interview %>%
  filter(person_id == "16371" | person_id == "14887")

# view(witness_interviews)

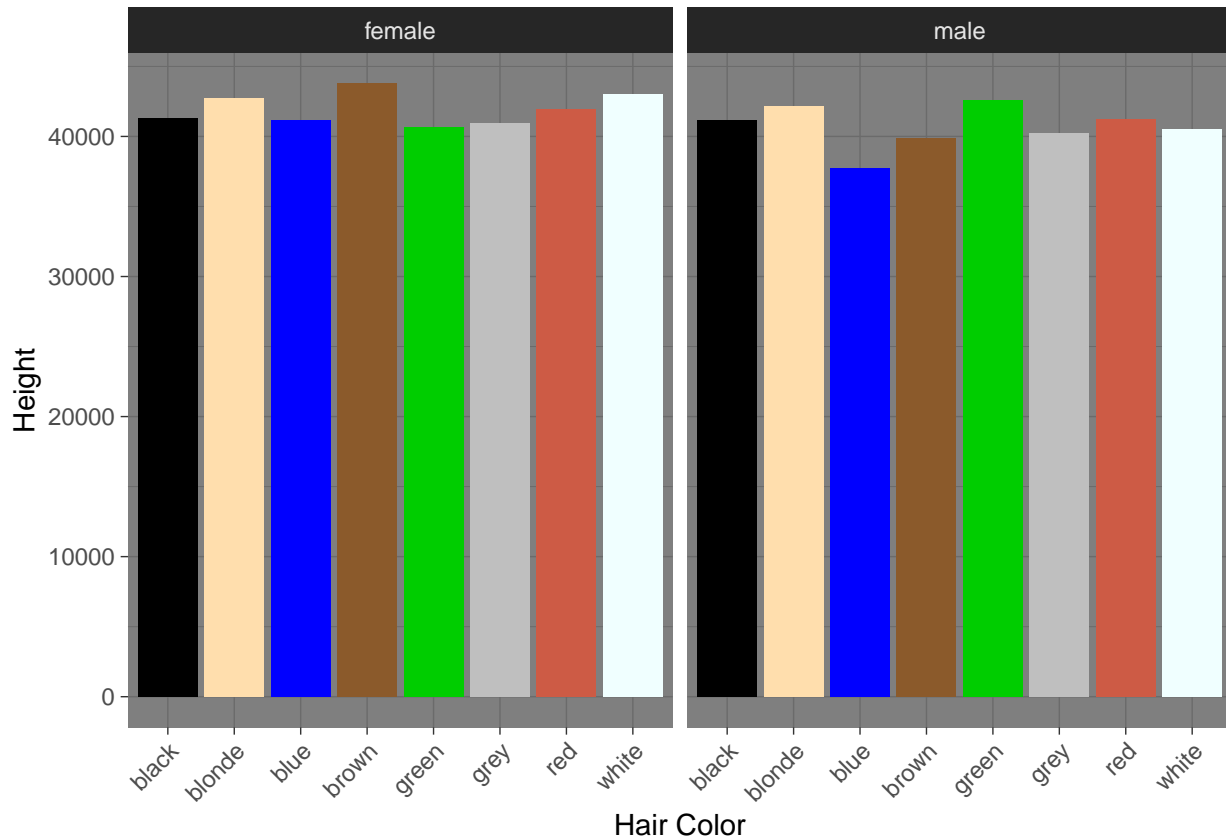
witness_interviews$transcript

## [1] "I heard a gunshot and then saw a man run out. He had a \"Get Fit Now Gym\" bag. The membership r
## [2] "I saw the murder happen, and I recognized the killer from my gym when I was working out last we
```

9. Create a single plot that compares the distribution of height across genders and hair colors (you can use any distribution plot type you like). Do you think this is real data or a toy dataset? Explain.

Toy dataset because of the relatively high counts of blue and green hair.

```
ggplot(drivers_license, aes(x=hair_color, y=height, fill=hair_color)) +
  geom_col() +
  scale_fill_manual(values = c(black="black", grey="grey75", white="azure", red="coral3", green="green3", blue="blue", brown="brown", orange="orange")) +
  facet_wrap(~ gender) +
  labs(x="Hair Color", y="Height") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position = "none")
```



10. What is the name of the murderer?

```
library(stringr)

murderer_file <- get_fit_now_member %>%
  filter(str_detect(id, "48Z") & membership_status == "gold")

# view(murderer_file)

plate <- drivers_license %>%
  filter(str_detect(plate_number, "H42W"))

# view(plate)

murderer <- person %>%
  filter((id == "28819" | id == "67318") & (license_id == "183779" | license_id == "423327" | license_id == "423328"))

# view(murderer)
```

The murderer is Jeremy Bowers!

11. What is the name of the true villain behind this crime?

```
murderer_interview <- interview %>%  
  filter(person_id == "67318")  
  
# view(murderer_interview)  
  
murderer_interview$transcript
```

```
## [1] "I was hired by a woman with a lot of money. I don't know her name but I know she's around 5'5\"
```

```
mystery_woman <- drivers_license %>%  
  filter(car_make == "Tesla" & car_model == "Model S" & hair_color == "red" & gender == "female")  
  
# view(mystery_woman)
```

```
woman_id <- person %>%  
  filter(license_id == "202298" | license_id == "291182" | license_id == "918773")  
  
# view(woman_id)
```

```
villain <- facebook_event_checkin %>%  
  filter((person_id == "78881" | person_id == "90700" | person_id == "99716") & str_detect(date,"201712  
  
# view(villain)
```

The true villain is Miranda Priestly!