GR Crash Data Proposal STA 518

Luke Cadagin

10/15/2021

Load Libraries:

```
library(tidyverse)
library(lubridate)
library(sf)
library(osmdata)
library(here)
```

Read-In Crash Data:

```
crash_data <- read_csv(here::here("data", "CGR_Crash_Data2.csv"))
```

Grand Rapids Crash Data Project

Proposal

Overview:

On September 18th, 2021 I was rear-ended in a hit-and-run accident. This incident got me interested in exploring data related to traffic accidents in the state of Michigan.

Dr. Kapitula pointed me to a traffic accident dataset on the grdata website (https://grdata-grandrapids.opendata.arcgis.com/) that includes de-identified crash data for all reports in Grand Rapids from 2007 to 2017.

Unfortunately the .csv file downloaded from the grdata website was slightly too large to be stored in github (around 103MB). I removed the CITY, COUNTY, MDOTREG, RDUSRINVID, NONTRAFFIC, and FRAMEWORK fields from the .csv, before uploading it in R/github to cut down on size. These columns contained the same value for every single row and did not assist with analysis.

End Product:

The goal of my project is to create a report or website for the City of Grand Rapids that analyzes traffic risks and provides suggestions on how improvements could be made.

Traffic Risk and Questions could include:

Do Hit and Run accidents ocurr in higher or lower income zip codes? Would public outreach programs in those zip codes explaining Michigan no-fault laws reduce the frequecy of hit and runs?

Do a majority of accidents involving trains ocurr at crossings that use gates? Would including more gates for pedestrians and vehicles increase public safety? This is an intersting topic as there are trains that go straight through the Pew Campus everyday.

Do posted speed limits on a street play a large role in driver injuries or deaths? Would decreasing speed limits in heavily populated areas decrease crash injuries and deaths?

Figure 1: Alt text

Are acidents with deer a major concern in rural areas? Do fences on rural roads effectively reduce accidents involving deer or other animals?

I will need to look for additional data sets (such as average income levels in grandrapids zip codes) to assist me with my analysis.

Graphing:

I can use the sf and osmdata libraries to graph the location of accidents on a map of Grand Rapids. This will allow me to identify clusters and trends of accidents spread out around the city.

For example, below is a visualizations of all hit and run accidents that ocurred in 2017 between the hours of 10:00 PM and 12:00 AM:

```
hit_and_run_2017 <- crash_data %>%
  filter(HITANDRUN == "Yes", YEAR == 2017, HOUR >= 22)
```

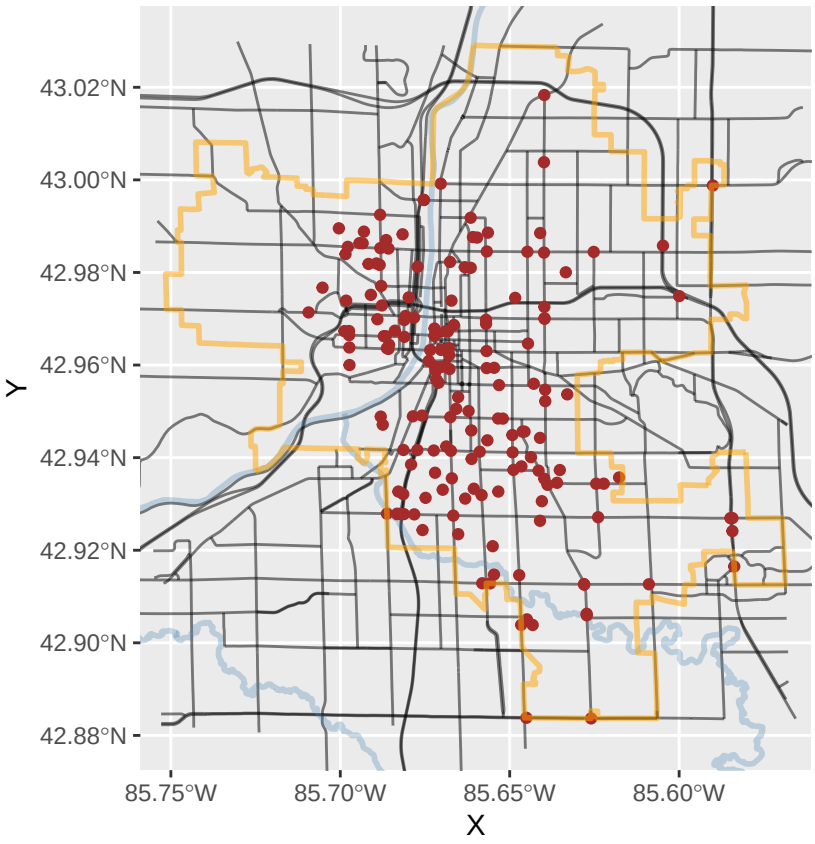
```
location_gr <- getbb("Grand Rapids") %>%
  opq()

roads_gr <- location_gr %>%
  add_osm_feature(key = "highway", value = c("motorway", "trunk", "primary", "secondary", "tertiary")) %
  osmdata_sf()

water_gr <- location_gr %>%
  add_osm_feature(key = "waterway", value = c("river")) %>%
  osmdata_sf()

boundary_gr <- location_gr %>%
  add_osm_feature(key = "boundary", value = "administrative") %>%
  #add_osm_feature(key = "admin_level", value = "8") %>%
  #add_osm_feature(key = "border_type", value = "city") %>%
  #add_osm_feature(key = "place", value = "city") %>%
  add_osm_feature(key = "name", value = "Grand Rapids") %>%
  osmdata_sf()

ggplot()+
         geom_sf(data = roads_gr$osm_lines, size = .5, alpha = .5, color = 'black') +
         geom_sf(data = water_gr$osm_lines, size = 1, alpha = .3, color = 'steelblue') +
         geom_point(data = hit_and_run_2017, mapping = aes(x = X, y = Y), color = "brown") +
         geom_sf(data = boundary_gr$osm_lines, size = 1, alpha = .5, color = "orange") +
         coord_sf(xlim = c(-85.57, -85.75), ylim = c(42.88, 43.03))
```



Challenges:

Data Entered Incorrectly:

Let's take a look at a Hit and Run that occurred on 6/16/2008 on Alger Street (Primary Road Name):

```
june_6_2008 <- crash_data %>%
  select(CRASHDATE, HOUR, HITANDRUN, DRIVER1AGE, DRIVER1SEX, DRIVER2AGE, DRIVER2SEX, PRNAME, INTERNAME)
  filter(CRASHDATE == "6/16/2008", HITANDRUN == "Yes", PRNAME == "ALGER")

june_6_2008
```

```
## # A tibble: 1 x 9
##   CRASHDATE  HOUR HITANDRUN DRIVER1AGE DRIVER1SEX DRIVER2AGE DRIVER2SEX PRNAME
##   <chr>     <dbl> <chr>          <dbl> <chr>           <dbl> <chr>      <chr>
## 1 6/16/2008    22 Yes               16 F                  18 M          ALGER
## # ... with 1 more variable: INTERNAME <chr>
```

You can see that the age of the vehicle one driver (in this instance, the victim of the hit and run) is entered as 16.

We can use the Michigan Traffic Crash Facts (MTCF) site to view a redacted police report of this same crash:

Here you can see that the driver was born 8/8/81, which would make them 26 at the time. We know this is the driver because they are located in the "UNIT/DRIVER" section on the left hand side. Also, position 1 is the driver position in a vehicle (as seen in the "UD-10 Traffic Crash Report Instruction Manual - 2018"):

However, there was a passenger in vehicle one who was born 7/25/91, which would make them 16 at the time. We know this is the passenger because they are located in the "PASSENGERS" section on the left hand side. Also, position 4 is not the driver position:

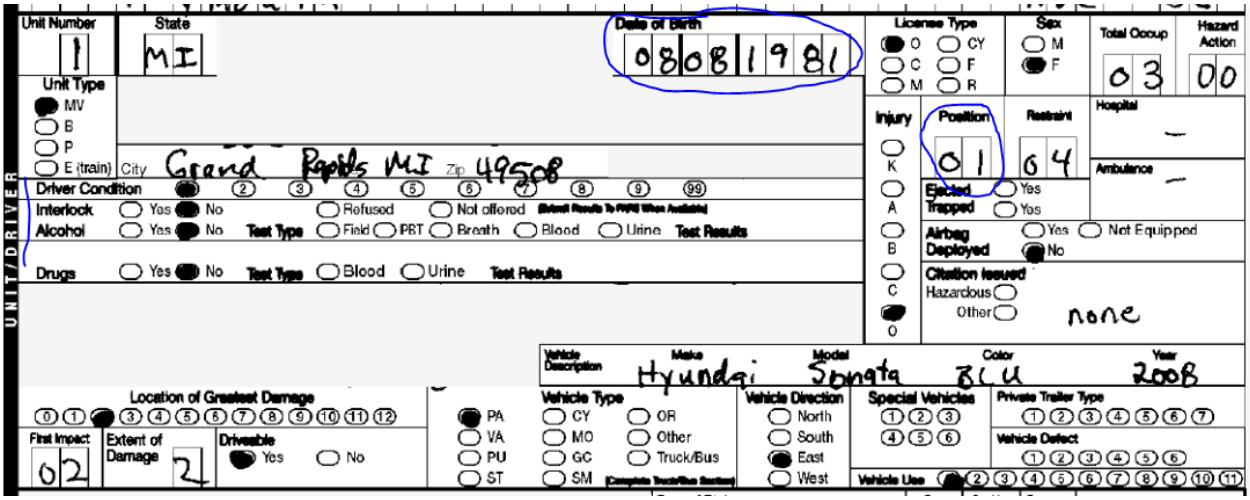


Figure 2: Alt text

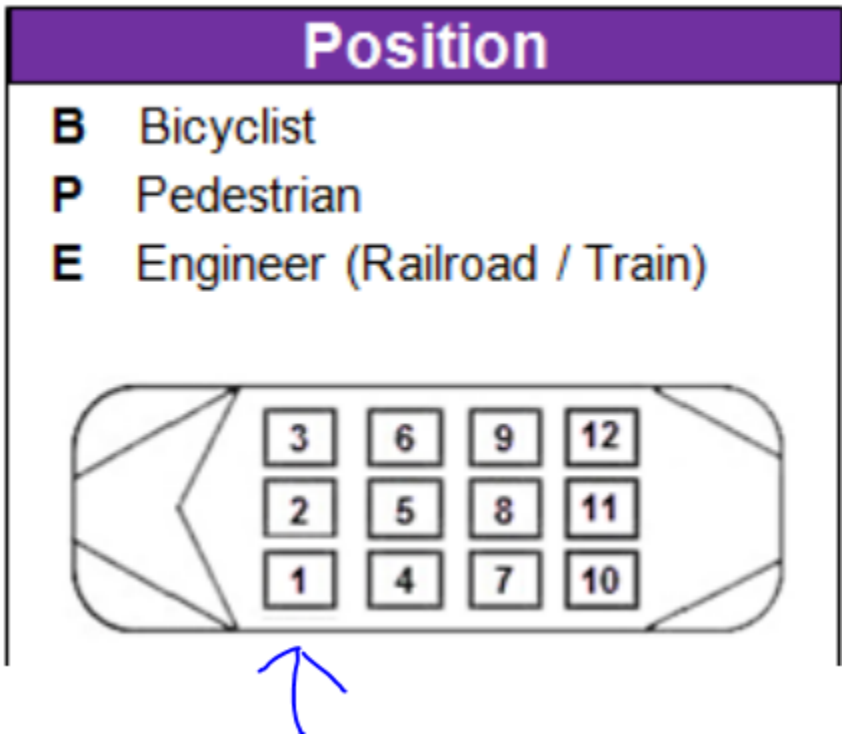


Figure 3: Alt text

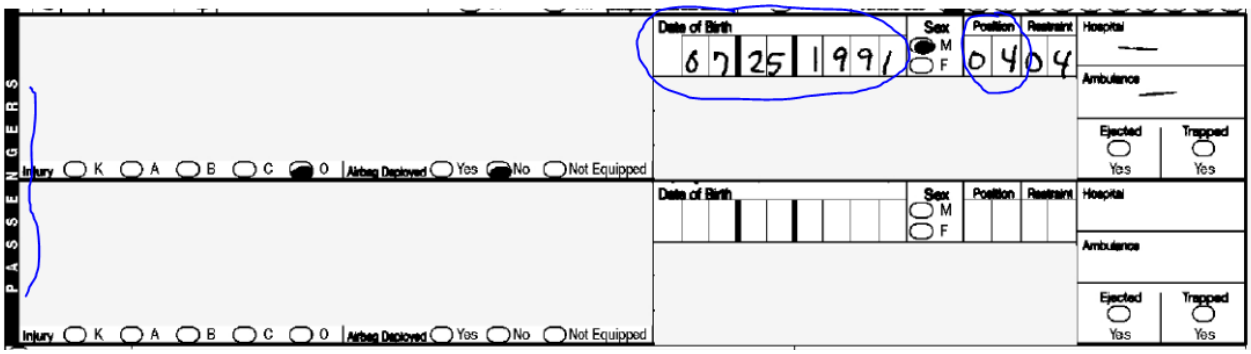


Figure 4: Alt text

This suggests that there some of the data entered into the csv file I downloaded from the grdata website was not correct.

It is interesting to note that the MTCF site also identities the driver of vehicle one as being 16 years old. This suggests that there is an error in the State of Michigan crash database. This was most likely an entry error when the hand-written report was digitized:

City or Township	Crash Day	Crash Month	Crash Year	Crash: Hit-and-Run	Crash: Young Driver	Total Units Reported	Time of Day	Lighting Conditions
Kent County: Grand Rapids	16	June	2008	Hit-and-Run	No Driver Age 15-24	2	2:00 PM - 2:59 PM	Daylight
Kent County: Grand Rapids	16	June	2008	Hit-and-Run	No Driver Age 15-24	1	12:00 Midnight - 12:59 AM	Dark - Lighted
Kent County: Grand Rapids	16	June	2008	Hit-and-Run	No Driver Age 15-24	2	1:00 PM - 1:59 PM	Other/Unknown
Kent County: Grand Rapids	16	June	2008	Hit-and-Run	Driver Age 16	2	10:00 PM - 10:59 PM	Dark - Lighted

Figure 5: Alt text

We can see that there are other potential errors present in the data:

```
under_5 <- crash_data %>%
  select(DRIVER1AGE) %>%
  filter(DRIVER1AGE < 5)

head(under_5)
```

```
## # A tibble: 6 x 1
##   DRIVER1AGE
##        <dbl>
## 1          2
## 2          3
## 3          0
## 4          0
## 5          4
## 6          1
```

A challenge of exploring this data will be to see if additional entry errors like this occurred and making adjustments to correct or mitigate these errors.

Generalizing the Data:

It is important to keep in mind that this dataset only spans from 2007 to 2017. We are not analyzing any data from 2018 to 2021, thus some driving trends or habits may have changed over time. For example, crash data after March 2020 is most likely atypical of prior time periods due to the COVID-19 lockdowns. In addition to this, we must be mindful that not all traffic accidents are reported to the police and our dataset most likely does not encompass the true population of crashes between 2007 and 2017.