Data Science Project Proposal:
LA Traffic-related Fatalities, 2009-2013

Elena Castellanos

Sally Rong

Motivation

Living in Southern California, both of us have grown up in a car. Cars are quintessential to being an Angeleno. So much so that, according to LA Weekly, there are more cars than people living in Los Angeles. Mind that we did say living as a verb to describe both cars and people. Los Angeles is a car-centric city. The city is quite literally built around the car – massive freeways and extra-wide streets frame the cityscape. Yet, the features that make Los Angeles ideal for a car are not necessarily ideal for people. In the city of Los Angeles alone, there were 240 traffic-related deaths in 2018 (LACurbed). Bicyclists and pedestrians are most impacted by this issue. Streets lacking crosswalks, bike lanes, sidewalks, or protected left turns can make even trying to use another form of transportation a safety hazard.

Understanding the severity of traffic-related deaths in Los Angeles, Mayor Eric Garcetti began the Vision Zero Initiative in 2015. Vision Zero is a data-driven road safety policy that aims to "eliminate all traffic fatalities and severe injuries, while increasing safe, healthy, equitable mobility for all." Vision Zero is Los Angeles' commitment to eliminate traffic deaths by 2025. Still, the city has mixed results with traffic fatalities still on the rise and 2025 quickly approaching.

Related Work

We were inspired by Ben Willington's parking spot ticketing example during his TED Talk focused on "Making data mean more through storytelling." His data analysis served as a framework for municipal action. Similarly, Vision Zero uses traffic-related data to prioritize municipal action based on the severity of traffic fatalities on a given street. Elena found out about Vision Zero two years ago, when she job-shadowed the Director of Los Angeles's Vision Zero Initiative. The initiative's impact on Los Angeles is best exemplified by its Safe Streets Project that includes events like CicLAvia (a biking festival that closes off streets of LA for bicyclists) and Play Street (a rotating monthly event that brings playground equipment to traffic-fatality stricken areas of Los Angeles). Below,we have attached links to blogs that have spoken on the impact or lack thereof, of the Vision Zero Initiative in Los Angeles.

https://la.streetsblog.org/2019/03/06/l-a-county-seeks-input-on-draft-vision-zero-action-plan/

https://laist.com/2020/04/21/la_city_budget_cuts_vision_zero_street_improvements.php

https://law888.com/los-angeles-vision-zero-plan-not-working-yet/

https://usa.streetsblog.org/2020/06/24/centering-equity-is-a-matter-of-life-and-death-responding-to-anti-black-racism-in-urbanism/

Data

The data we are looking at is collected SWITRS (Statewide Integrated Traffic Records System) Data from 2012-2016, and accounts for collisions that occurred during that time frame within the City of Los Angeles' jurisdiction. The dataset is available on the Los Angeles GeoHub, the City's public platform for exploring and downloading location-based Open Data. The SWITRS dataset is of fairly large size (90 columns and 171,534 rows) and contains information on an abundance of various attributes associated with traffic collisions, allowing for more levels of analysis. Analysis of traffic collisions in Los Angeles city can be conducted on, for example, the collision level, party level (driver, pedestrian, bicyclist, etc.), and/or victim levels. Looking at the dataset currently, it appears clean and ready to be analyzed, and we do not anticipate much data cleaning needed, if any at all. Los Angeles's Vision Zero Initiative uses the detailed collision data to create a "collision landscape" that is foundational to their data analysis and street modifications. The SWITRS data set is also used by the Initiative to establish and maintain a High Injury Network (HIN), which comprises streets with a high concentration of either serious or fatal traffic collisions.

Questions

Interestingly, the High Injury Network (HIN) represents only about 6% of LA city's streets (over 450 miles), yet accounts for over 65% of severe or fatal traffic injuries for people walking. A potential engaging question to pose here is whether there are recurring factors that contribute to traffic collisions at HIN streets and what those factors are. For instance, are there certain modes of transportation (example: walking vs. biking) more frequently associated with severe or fatal traffic collisions? Are certain community areas within the city of Los Angeles more affected by fatal or traffic collisions than others? Are there certain driving movements/behaviors (changing lanes, parking, turns, etc.) that contribute more or less to traffic accidents? We are unsure whether the SWITRS dataset contains sufficient data to answer these questions, but exploring the extent of the relationship between the density of severe/fatal traffic collisions and a communities' health outcomes is also an interesting question to pose and give consideration to using this dataset.

Possible Findings and Implications

After conducting analysis of the data, we hope to make informed recommendations for executing the Vision Zero Initiative in the City of Los Angeles. By identifying factors that

contribute to a high number of fatal or severe traffic collisions in concentrated streets or intersections, LA City can begin to prioritize safety improvements in those areas. In addition, the High Injury Network (HIN), prioritized locations can be identified for upcoming safety projects and to develop “countermeasure pairing,” which is the process of pinpointing architectural and engineering countermeasures to adequately address collisions that have similar attributes. Together, these datasets are used by Vision Zero transportation planners to distinguish “Priority Corridors,” which are streets that are a minimum of 0.5 miles long and have witnessed at least 15 traffic deaths and serious injuries per mile over the 2013-2017 five year period. The data collected and analyzed to identify priority corridors is what leads to municipal action. Vision Zero improvements can be as seemingly small as a Leading Pedestrian Interval, which gives pedestrians a 3-7 second head start to cross an intersection. By analyzing the SWITRS dataset and creating safety project recommendations for LA City, we are meeting the Vision Zero Initiative aims of eliminating preventable traffic-related deaths.A data-driven approach can integrate lessons learned from human failing and move forward to creating policies that will help eliminate fatal and severe traffic accidents. Los Angeles’s commitment to safer streets has the potential to make Los Angeles a city built for people not for cars.

2009-2013 LA Traffic-related Fatalities & Attributes: A Surface Level Examination

Sally Rong, Elena Castellanos

08/06/2020

Presentation Slides

YouTube Presentation

SWITRS 2009-2013 Collisions Dataset

Victims Table

Party Table

Codebook

The codebook provides analysis on 3 levels: the party level, victim level, and collision level.

#rm(list=ls()) #setwd("~/Desktop/Data Science") #install.packages('tidyverse')

Load the data

```
library(tidyverse)
```

```
## ── Attaching packages ──────────────────────────────────────── tidyverse 1.3.
0 ──
```

```
## ✓ ggplot2 3.3.0     ✓ purrr   0.3.3
## ✓ tibble  3.0.3     ✓ dplyr   0.8.3
## ✓ tidyr   1.1.1     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.4.0
```

```
## Warning: package 'tibble' was built under R version 3.6.2
```

```
## Warning: package 'tidyr' was built under R version 3.6.2
```

```
## ── Conflicts ──────────────────────────────────────── tidyverse_conflicts
() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
collisions2 <- read_csv('Collisions_2009-2013_(SWITRS).csv')
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   X = col_double(),
##   Y = col_double(),
##   FID = col_double(),
##   CASE_ID = col_double(),
##   ACCIDENT_YEAR = col_double(),
##   JURIS = col_double(),
##   COLLISION_TIME = col_double(),
##   DAY_OF_WEEK = col_double(),
##   SHIFT = col_double(),
##   POPULATION = col_double(),
##   CNTY_CITY_LOC = col_double(),
##   SPECIAL_COND = col_double(),
##   BEAT_TYPE = col_double(),
##   CHP_BEAT_CLASS = col_double(),
##   DISTANCE = col_double(),
##   CALTRANS_DISTRICT = col_double(),
##   STATE_ROUTE = col_double(),
##   POSTMILE = col_double(),
##   COLLISION_SEVERITY = col_double(),
##   NUMBER_KILLED = col_double()
##   # ... with 19 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
party_table <- read_csv('Party_Tables_-_Collisions_2009-2013_(SWITRS).csv')
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   FID = col_double(),
##   PARTY_NUMBER = col_double(),
##   PARTY_TYPE = col_double(),
##   PARTY_AGE = col_double(),
##   OAF_VIOL_SECTION = col_double(),
##   PARTY_NUMBER_KILLED = col_double(),
##   PARTY_NUMBER_INJURED = col_double(),
##   VEHICLE_YEAR = col_double(),
##   ACCIDENT_YEAR = col_double()
## )
## See spec(...) for full column specifications.
```

```
victim_table <- read_csv('Victim_Tables_-_Collisions_2009-2013_(SWITRS).csv')
```

```
## Parsed with column specification:
## cols(
##   FID = col_double(),
##   CASE_ID = col_character(),
##   PARTY_NUMBER = col_double(),
##   VICTIM_ROLE = col_double(),
##   VICTIM_SEX = col_character(),
##   VICTIM_AGE = col_double(),
##   VICTIM_DEGREE_OF_INJURY = col_double(),
##   VICTIM_SEATING_POSITION = col_double(),
##   VICTIM_SAFETY_EQUIP_1 = col_character(),
##   VICTIM_SAFETY_EQUIP_2 = col_character(),
##   VICTIM_EJECTED = col_double(),
##   ACCIDENT_YEAR = col_double()
## )
```

PARTY-VICTIM relationship

Merge party_table & victim_table, by CASE ID and any number of party records that can be associated w/ a collision

```
party_victim <- left_join(x=party_table, y=victim_table, by= c("CASE_ID", "PARTY_NUMBER"))
dim(party_victim)
```

```
## [1] 1110455      45
```

COLLISION-PARTY relationship

For this join, we faced difficulting tying the two datasets by CASE_ID. We have ran this problem through with Professor Raja too.

```
#collision_party <- left_join(x=collisions2, y=party_table, by= c("CASE_ID"))
```

```
df <- party_victim %>%
  select(PARTY_AGE, PARTY_SEX, VICTIM_DEGREE_OF_INJURY, VICTIM_SAFETY_EQUIP_1, VICTIM_ROLE, VICTIM_AGE, VEHICLE_Y
EAR, VICTIM_SEX, STWD_VEHICLE_TYPE, VEHICLE_MAKE)
dim(df)
```

```
## [1] 1110455      10
```

We first looked only at fatal accidents.

```
fatal <- filter(df, VICTIM_DEGREE_OF_INJURY == 4)
dim(fatal)
```

```
## [1] 277934     10
```

```
fatal %>%
  group_by(VICTIM_SAFETY_EQUIP_1) %>%
  summarize(num = n()) %>%
  arrange(desc(num)) %>%
  mutate(percentage = num/277934)
```

VICTIM_SAFETY_EQUIP_1 <chr>	num <int>	percentage <dbl>
M	128730	4.631675e-01
L	51393	1.849108e-01
P	29836	1.073492e-01
-	27727	9.976109e-02
G	26889	9.674599e-02
N	5541	1.993639e-02
V	1367	4.918434e-03
A	1350	4.857268e-03
Y	1220	4.389531e-03
B	862	3.101456e-03

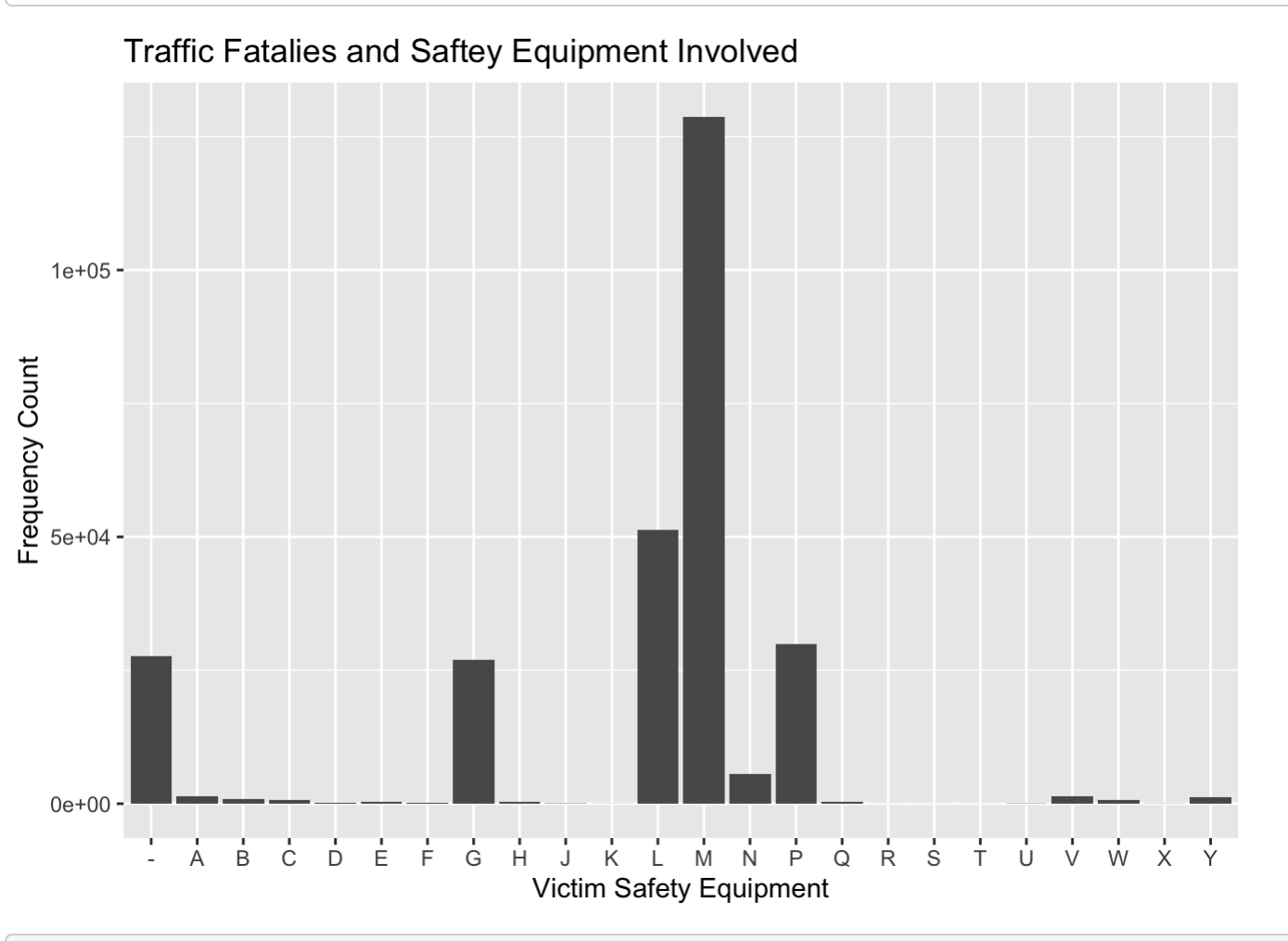
1-10 of 24 rows Previous 1 2 3 Next

It appears that in accidents that resulted in the victim killed, about 46.3% of those accidents involved the air bag not deployed ('M' above).

Here's a graph showing the Victim Safety Equipment used in fatal traffic accidents. Most frequently occuring is the air bag not deployed (M).

```
fatal %>%
  ggplot(mapping = aes(x=VICTIM_SAFETY_EQUIP_1)) +
  geom_histogram(
    stat = 'count') +
    labs(y = 'Frequency Count',
          x= 'Victim Safety Equipment',
        title = 'Traffic Fatalies and Saftey Equipment Involved')
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```
no_air_bag_fatal <- filter(df, VICTIM_SAFETY_EQUIP_1 == 'M', VICTIM_DEGREE_OF_INJURY == 4)
dim(no_air_bag_fatal)
```

```
## [1] 128730     10
```

128730/277934 of traffic accidents in which victim was killed involved air bag not deployed

We then looked at the most frequent ages of parties involved in fatal accidents in which the air bag was not deployed.

```
no_air_bag_fatal %>%
  filter(PARTY_AGE != 998) %>%
  group_by(PARTY_AGE) %>%
  summarize(number_parties_in_fatal_accidents = n()) %>%
  arrange(desc(number_parties_in_fatal_accidents))
```

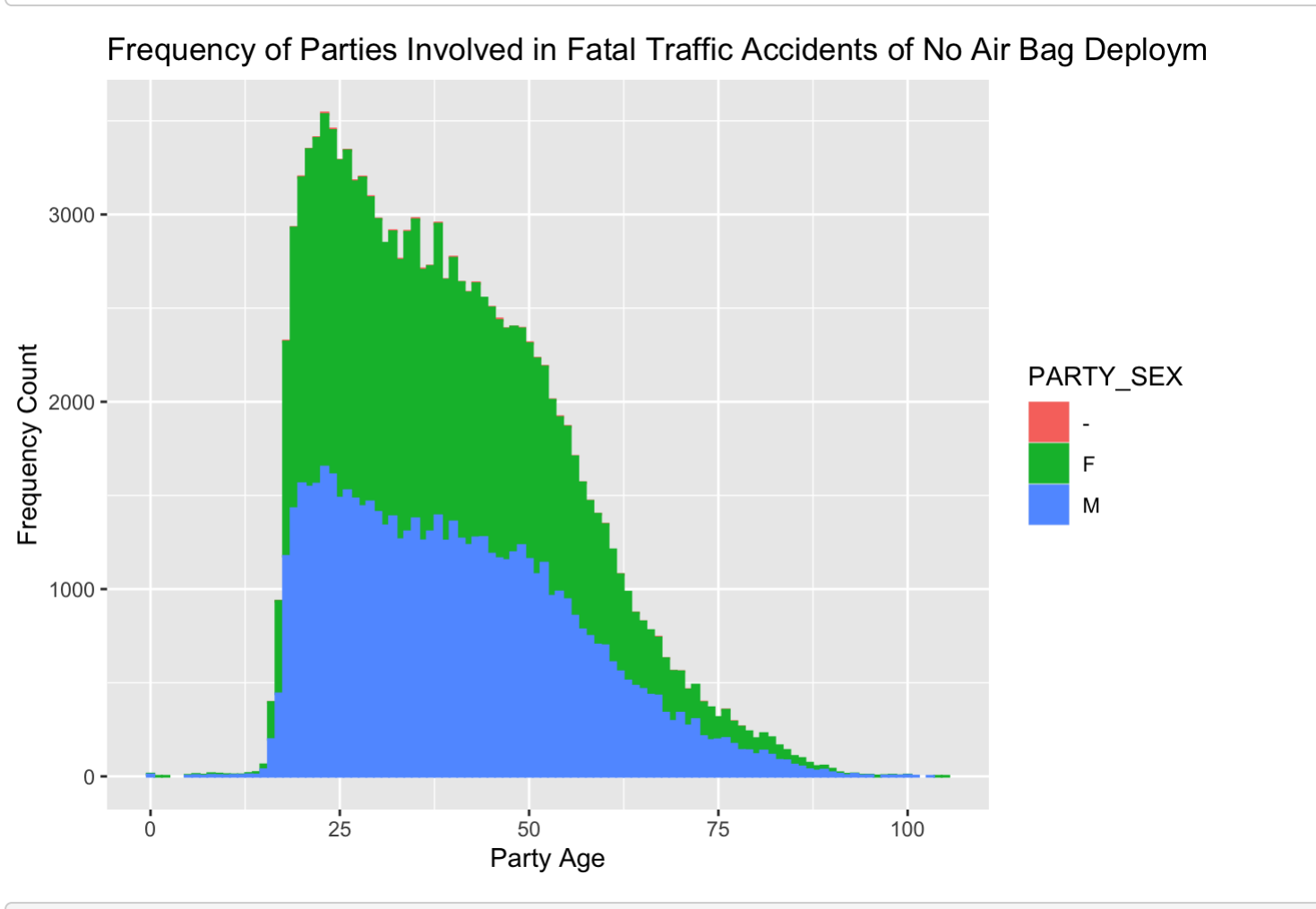
PARTY_AGE <dbl>	number_parties_in_fatal_accidents <int>
23	3543
24	3457
22	3411
21	3349
26	3344
25	3290
20	3202
28	3201
27	3181
29	3096

1-10 of 103 rows Previous 1 2 3 4 5 6 ... 11 Next

Younger parties in their twenties are more frequent in fatal traffic accidents of no air bag deployment.

```
no_air_bag_fatal %>%
  filter(PARTY_AGE != 998 ) %>%
  ggplot(mapping = aes(x=PARTY_AGE, fill = PARTY_SEX, color= PARTY_SEX)) +
  geom_histogram(
    stat = "count",
    bin = 0.5) +
  labs(y = 'Frequency Count',
       x= 'Party Age',
        title = 'Frequency of Parties Involved in Fatal Traffic Accidents of No Air Bag Deployment by Age')
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad, bin
```




```
filter (no_air_bag_fatal) %>%
  group_by(VEHICLE_MAKE) %>%
  summarize(num = n()) %>%
  arrange(desc(num))
```

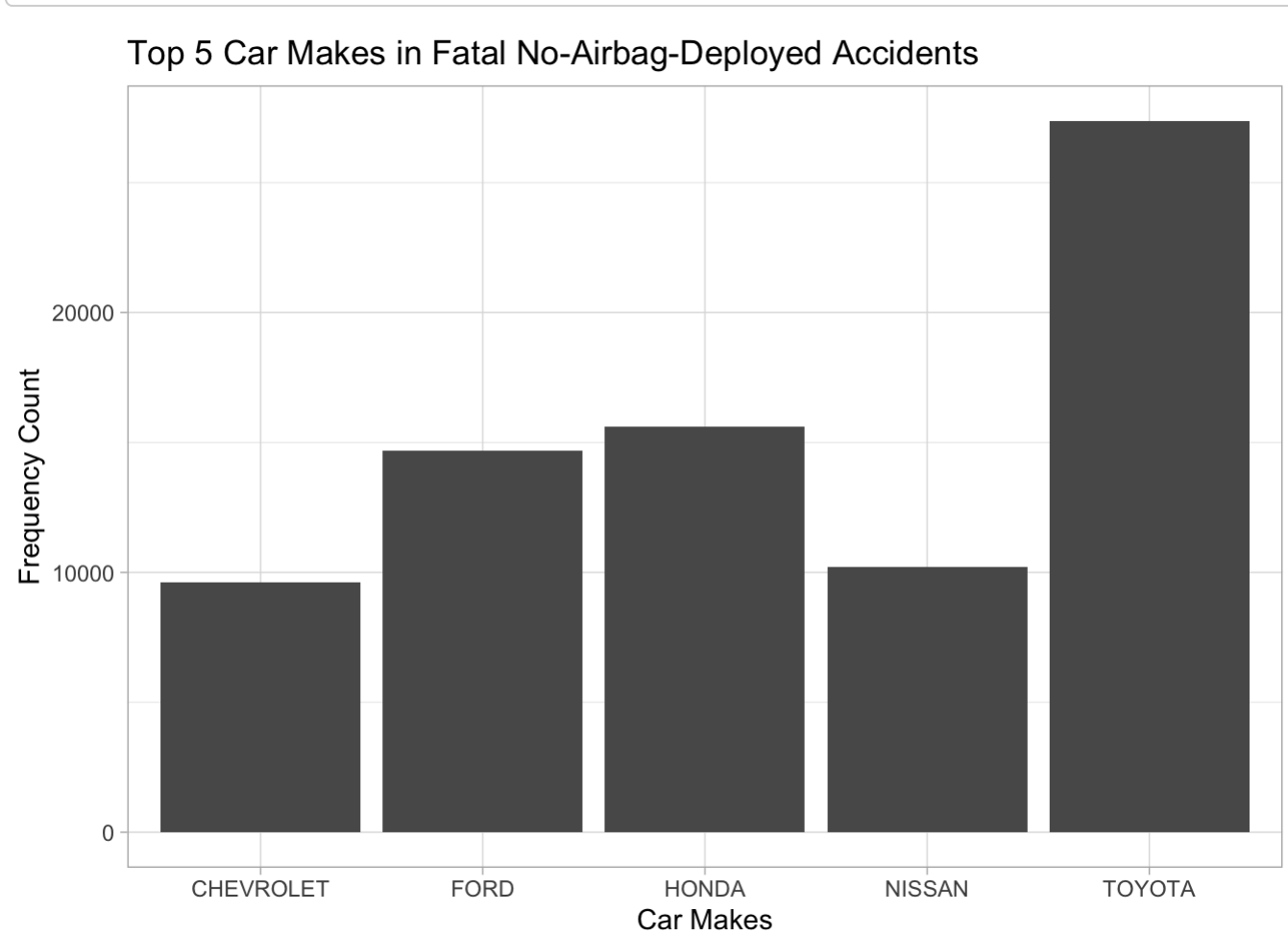
VEHICLE_MAKE <chr>	num <int>
TOYOTA	27376
HONDA	15593
FORD	14692
NISSAN	10201
CHEVROLET	9632
MERCEDES-BENZ	4983
DODGE	4390
BMW	3852
LEXUS	3682
CHRYSLER	2519

1-10 of 111 rows Previous 1 2 3 4 5 6 ... 12 Next

Of the fatal accidents without the airbag deployed, these are the top car makes.

```
no_air_bag_fatal %>%
  filter( VEHICLE_MAKE %in% c('TOYOTA', 'HONDA', 'FORD',  'NISSAN', 'CHEVROLET')) %>%
  ggplot(mapping = aes(x=VEHICLE_MAKE, fill = VEHICLE_YEAR, color= VEHICLE_YEAR)) +
  geom_histogram(
    stat = 'count') +
    theme_light() +
    labs(y = 'Frequency Count',
          x= 'Car Makes',
        title = 'Top 5 Car Makes in Fatal No-Airbag-Deployed Accidents')
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Interestingly, some car companies had to recall their cars for defective airbags (more info in the links below). The companies with most cars with the defective Takata airbags were also Honda, Ford, and Toyota. https://thomasjhenrylaw.com/blog/product-liability/takata-airbag-recall-a-complete-list-of-affected-vehicles/ https://www.consumerreports.org/car-recalls-defects/takata-airbag-recall-everything-you-need-to-know/] https://www.newsweek.com/airbag-recall-list-car-models-toyota-ford-honda-bmw-nissan-takata-1483468

```
no_air_bag_fatal %>%
  group_by(VEHICLE_YEAR) %>%
  summarize(num = n()) %>%
  arrange(desc(num))
```

VEHICLE_YEAR <dbl>	num <int>
2003	8746
2002	8554
2000	8544
2005	8469
2001	8434
2004	8376
2006	7109
1999	6994
2007	6665
1998	5786

1-10 of 68 rows Previous 1 2 3 4 5 6 7 Next

Of Vehicles involved in fatal accidents of no air bag deployment, these are the top car years of vehicles involved. This dataset doesn't include information on specific car models, though the articles do point out the top car make and models involved in the Takata air bag recall.

Now we will examine fatal accidents including bicyclists only.

```
bicyclist_fatal <- filter(df, VICTIM_ROLE == 4 & VICTIM_DEGREE_OF_INJURY == 4)
dim(bicyclist_fatal)
```

```
## [1] 8869   10
```

```
bicyclist_fatal %>%
  filter(VICTIM_SAFETY_EQUIP_1 != '-' ) %>%
  group_by(VICTIM_SAFETY_EQUIP_1) %>%
  summarize(num = n()) %>%
  arrange(desc(num))
```

VICTIM_SAFETY_EQUIP_1 <chr>	num <int>
P	3925
V	1268
W	237
N	107
M	102
A	99
B	19
L	8
X	6
Y	5

1-10 of 17 rows Previous 1 2 Next

Most frequent was P - not required. We didn't evaluate that, since the codebook wasn't clear on what that means. We looked at V instead, the 2nd most frequent - 'Driver, Motorcycle Helmet not Used.' We also included X - 'Passenger, motorcycle, Helmet not used' for a more robust look at cases that the helmet was not used.

```
bicyclist_fatal %>%
  filter(VICTIM_SAFETY_EQUIP_1 %in% c('V', 'X')) %>%
  group_by(VICTIM_AGE) %>%
  summarize(num = n()) %>%
  arrange(desc(num))
```

VICTIM_AGE <dbl>	num <int>
20	62
18	50
19	50
22	41
16	39
30	39
21	38
27	38
23	37
25	36

1-10 of 70 rows Previous 1 2 3 4 5 6 7 Next

Again, it seems bicyclists in at very young ages, 18-20 years, appear the most frequent in these fatal accidents in which helmets aren't used. Bicyclists over the age of 18 are not required to wear helmets. Perhaps deeper analysis may provide information whether there is correlation between these two factors.

https://www.bicyclelaw.com/bicycle-laws/california-bicycle-laws/california-bicycle-helmet-law/

```
bicyclist_fatal %>%
  filter(VICTIM_SAFETY_EQUIP_1 %in% c('V', 'X') & VICTIM_AGE != 998) %>%
  ggplot(mapping = aes(x=VICTIM_AGE, fill = VICTIM_SEX, color= VICTIM_SEX)) +
  geom_histogram(
    stat = "count",
    bin = 0.5) +
  labs(y = 'Frequency Count',
       x= 'Bicyclist Age',
        title = 'Frequency of Bicyclist Victims Involved in Fatal Traffic Accidents without Helmet')
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad, bin
```