DATA 606 Data Project Proposal

Data Preparation

```
# Imports
library(dplyr)
library(ggplot2)

# Set up our url
crash_url <- "https://raw.githubusercontent.com/riverar9/cuny-msds/main/data606/project/01_proposal/Cras

# Read in the url using read.csv
crash_df <- read.csv(
  crash_url
)
# Inspect the raw dataframe
head(crash_df)

# Since the dataframe is large, let's filter it to what we're interested in
crash_df <- crash_df |>
  select(
    Vehicle.Body.Type,
    ACRS.Report.Type,
    Crash.Date.Time
  ) |>
  filter(
    Vehicle.Body.Type != ""
  )

crash_df |>
  group_by(ACRS.Report.Type) |>
  summarise(count = n())

total_report_count <- nrow(crash_df)
```

Research question

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

Is there a difference in the crash rate between cars of different body types for Fatal and Injury crashes?

Cases

What are the cases, and how many are there?

Each case represents one vehicle crash incident report. There are 169,635 observations in this dataset.

Data collection

Describe the method of data collection.

This data was retrieved from (https://data.gov/)[https://data.gov/]. This is a resource from the United States Federal Government that provides datasets for free use.

Type of study

What type of study is this (observational/experiment)?

This is an observaitonal study.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

This data was obtained from Data.gov.

(URL to the data used.)[https://catalog.data.gov/dataset/crash-reporting-drivers-data]

Dependent Variable

What is the response variable? Is it quantitative or qualitative?

Our Dependent variable will be the number of crashes and it is quantitative.

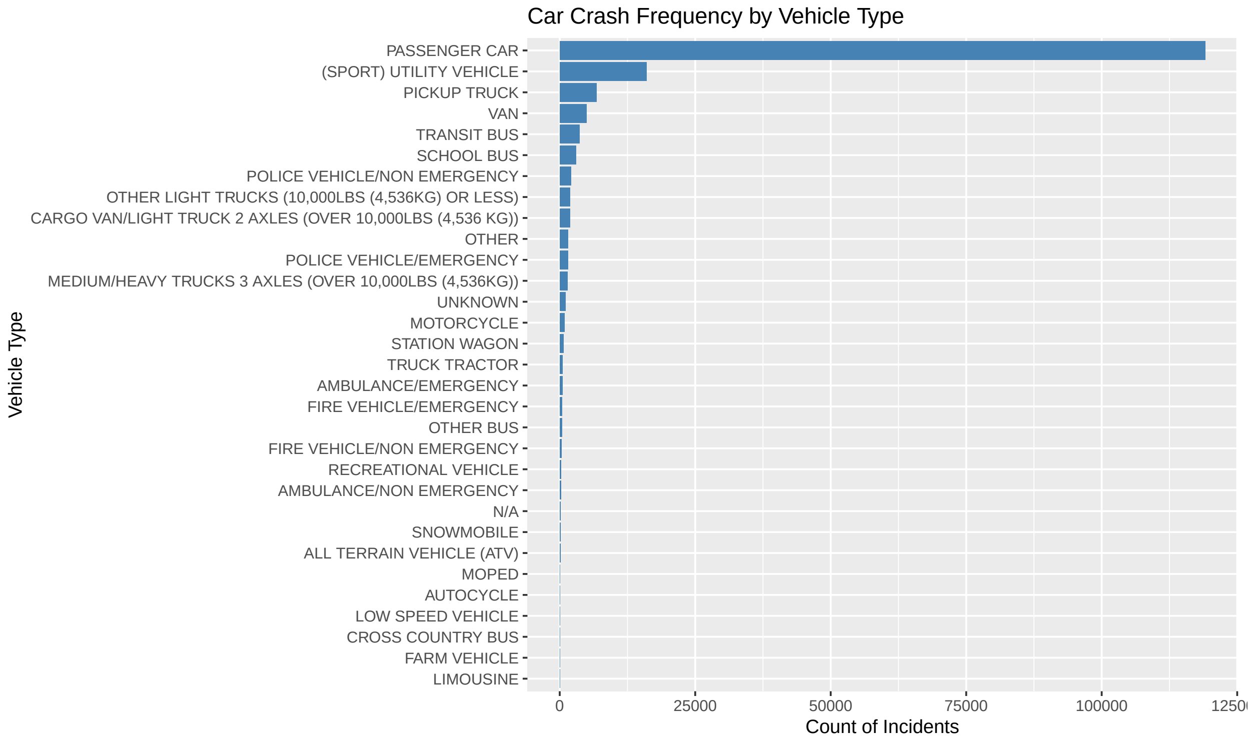
Independent Variable(s)

Our Independent variable will be: * Vehicle.Body.Type.

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

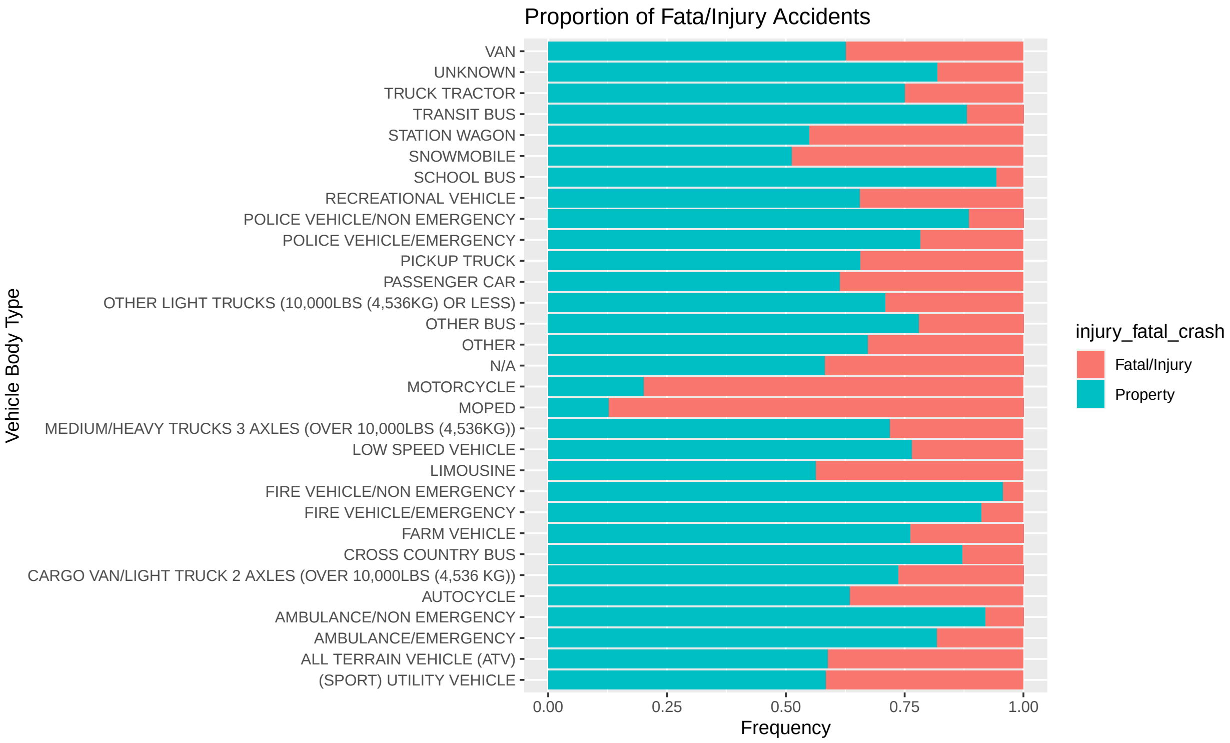
```
frequency_data <- crash_df |>
  count(Vehicle.Body.Type)

ggplot(
  frequency_data,
  aes(
    x = n,
    y = reorder(
      Vehicle.Body.Type,
      n
    )
  )
) +
  geom_bar(
    stat = "identity",
    fill = "steelblue"
  ) +
  labs(
    title = "Car Crash Frequency by Vehicle Type",
    x = "Count of Incidents",
    y = "Vehicle Type"
  )
```



```
vehicle_summary <- crash_df |>
  group_by(
    Vehicle.Body.Type
  ) |>
  summarise(
    vehicle_occurances = n()
  )

vehicle_crash_summary <- crash_df |>
  mutate(
    injury_fatal_crash = ifelse(
      ACRS.Report.Type == "Fatal Crash"
      | ACRS.Report.Type == "Injury Crash",
      "Fatal/Injury",
      "Property"
    )
  ) |>
  group_by(
    Vehicle.Body.Type,
    injury_fatal_crash
  ) |>
  summarise(
    vehicle_crash_occurances = n()
  )
```

```
## ‘summarise()‘ has grouped output by ’Vehicle.Body.Type’. You can override using
## the ‘.groups‘ argument.
```

```
vehicle_crash_summary <- left_join(
  vehicle_crash_summary,
  vehicle_summary,
  by = "Vehicle.Body.Type"
)

vehicle_crash_summary <- vehicle_crash_summary |>
  mutate(
    crash_freq = vehicle_crash_occurances / vehicle_occurances
  )

ggplot(
  data = vehicle_crash_summary,
  aes(
    x = crash_freq,
    y = reorder(
      Vehicle.Body.Type,
      crash_freq
    ),
    fill = injury_fatal_crash
  )
) +
  geom_bar(
    stat = 'identity'
  ) +
  labs(
    x = "Frequency",
    y = "Vehicle Body Type",
    title = "Proportion of Fata/Injury Accidents"
  )
```


Proportion of Fata/Injury Accidents
Vehicle Body Type
VAN
UNKNOWN
TRUCK TRACTOR
TRANSIT BUS
STATION WAGON
SNOWMOBILE
SCHOOL BUS
RECREATIONAL VEHICLE
POLICE VEHICLE/NON EMERGENCY
POLICE VEHICLE/EMERGENCY
PICKUP TRUCK
PASSENGER CAR
OTHER LIGHT TRUCKS (10,000LBS (4,536KG) OR LESS)
OTHER BUS
OTHER
N/A
MOTORCYCLE
MOPED
MEDIUM/HEAVY TRUCKS 3 AXLES (OVER 10,000LBS (4,536KG))
LOW SPEED VEHICLE
LIMOUSINE
FIRE VEHICLE/NON EMERGENCY
FIRE VEHICLE/EMERGENCY
FARM VEHICLE
CROSS COUNTRY BUS
CARGO VAN/LIGHT TRUCK 2 AXLES (OVER 10,000LBS (4,536 KG))
AUTOCYCLE
AMBULANCE/NON EMERGENCY
AMBULANCE/EMERGENCY
ALL TERRAIN VEHICLE (ATV)
(SPORT) UTILITY VEHICLE
0.00
0.25
0.50
0.75
1.00
Frequency
injury_fatal_crash
Fatal/Injury
Property