Introductions (3-4 sentences introducing the topic, citing relevant studies, etc.)

Road accidents are a serious and significant problem in the United States. Many Americans would say that they experience the fear that comes with driving. In fact, 66% of Americans reportedly experience driving anxiety (Meyer, 2023). When it comes to driving, drivers must pay close attention to the road and others around them for the safety of themselves and others. With driving comes the unpredictability of the weather, which can add to the anxiety of driving. Weather conditions such as rain, snow, and fog can impair vision while driving. This can cause dangerous accidents on the road as a result. We want to examine how different weather conditions (like rain, snow, and overcast), precipitation (inches), visibility (miles), wind speed (mph), and temperature (F) affect the severity of car accidents?

IVs: Weather condition, precipitation (inches), visibility (miles), wind speed (mph), temperature (F)
DVs: Severity (1 to 4)
https://smoosavi.org/datasets/us_accidents

Research question and hypothesis (1 sentence each)

RQ: How do different weather conditions (like rain, snow, and overcast), precipitation (inches), visibility (miles), wind speed (mph), and temperature (F) affect the severity of car accidents?

Hypothesis: The frequency of road accidents will increase with worsening weather conditions, and accident severity will increase with accident occurrences.

Potential Articles:

1. "The Relationship Between Road Accident Severity and Recorded Weather" https://www.sciencedirect.com/science/article/abs/pii/S0022437598000516?via%3Dihub
2. "Weather impacts on various types of road crashes: a quantitative analysis using generalized additive models" https://etrr.springeropen.com/articles/10.1186/s12544-022-00561-2
3. "The effects of weather factors on road traffic casualties: Analysis on provincial panel data of China from 2006 to 2021" https://www.cell.com/heliyon/fulltext/S2405-8440(24)12819-

8?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2405844024128198%3Fshowall%3Dtrue

4. "Good weather for a ride (or not?): how weather conditions impact road accidents — a case study from Wielkopolska (Poland)" https://pmc.ncbi.nlm.nih.gov/articles/PMC10794278/

Dataset: US Accidents (2016 - 2023) 3 GB of data, would need to be reduced in R, however, shows fatality, severity, multiple weather conditions

Description of your data (how did you get it, what does it contain, could use summary graphics, etc.)

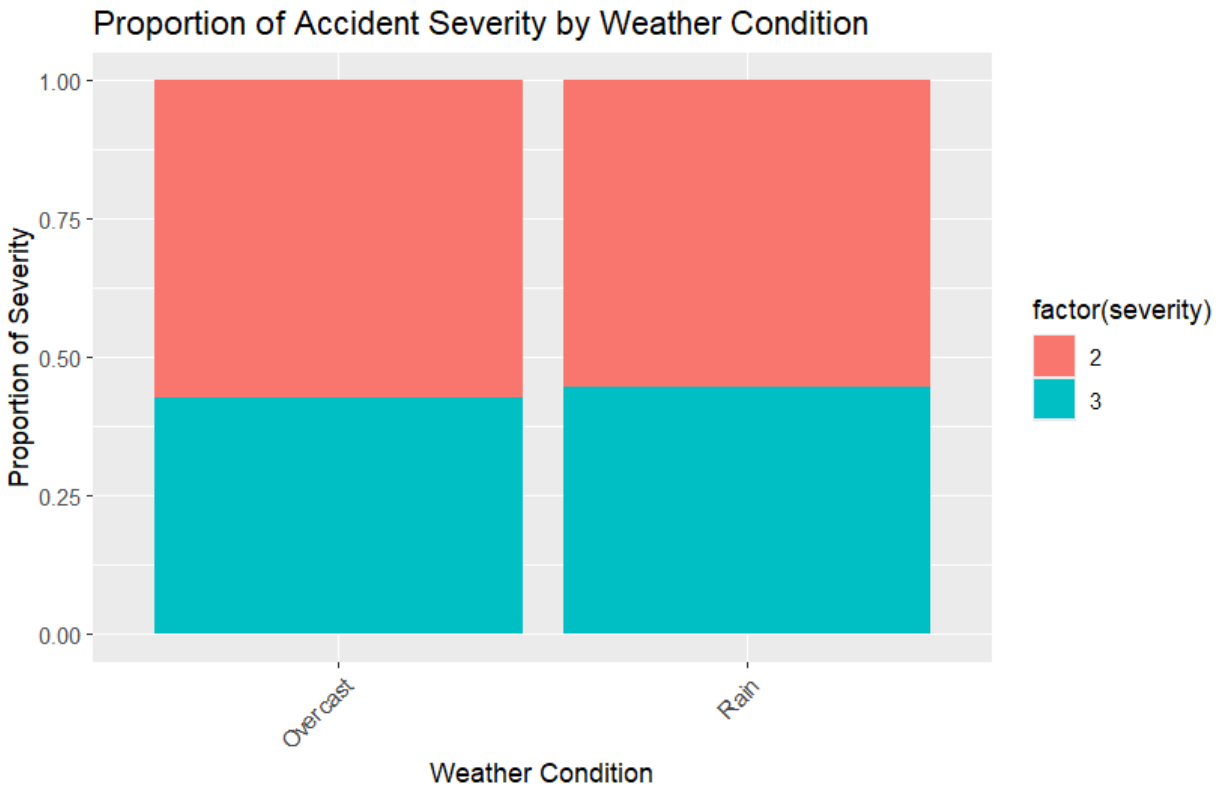
- Kaggle/Article link to dataset
- Summary
    - Head(dataset)
    - (optional but might help) Remove columns "start_time, end_time,end_lat, end_lng"
    - Str(dataset)
    - Missing data?
    - Outliers?
        - Simple R code or using cook's distance, leverage, or jacknife residuals
    - Rows
    - Columns
    - Mean, range, sd, quantile

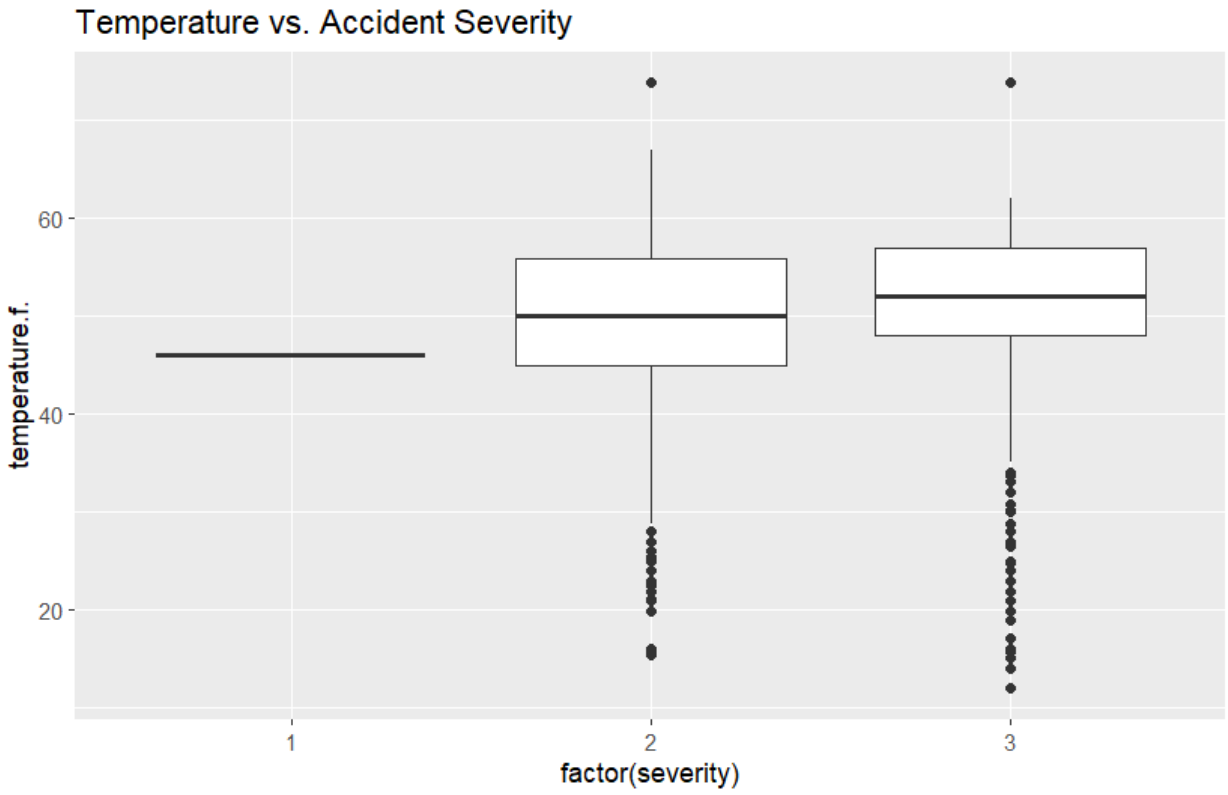
Analysis and Conditions/Assumptions (can be put under analysis)

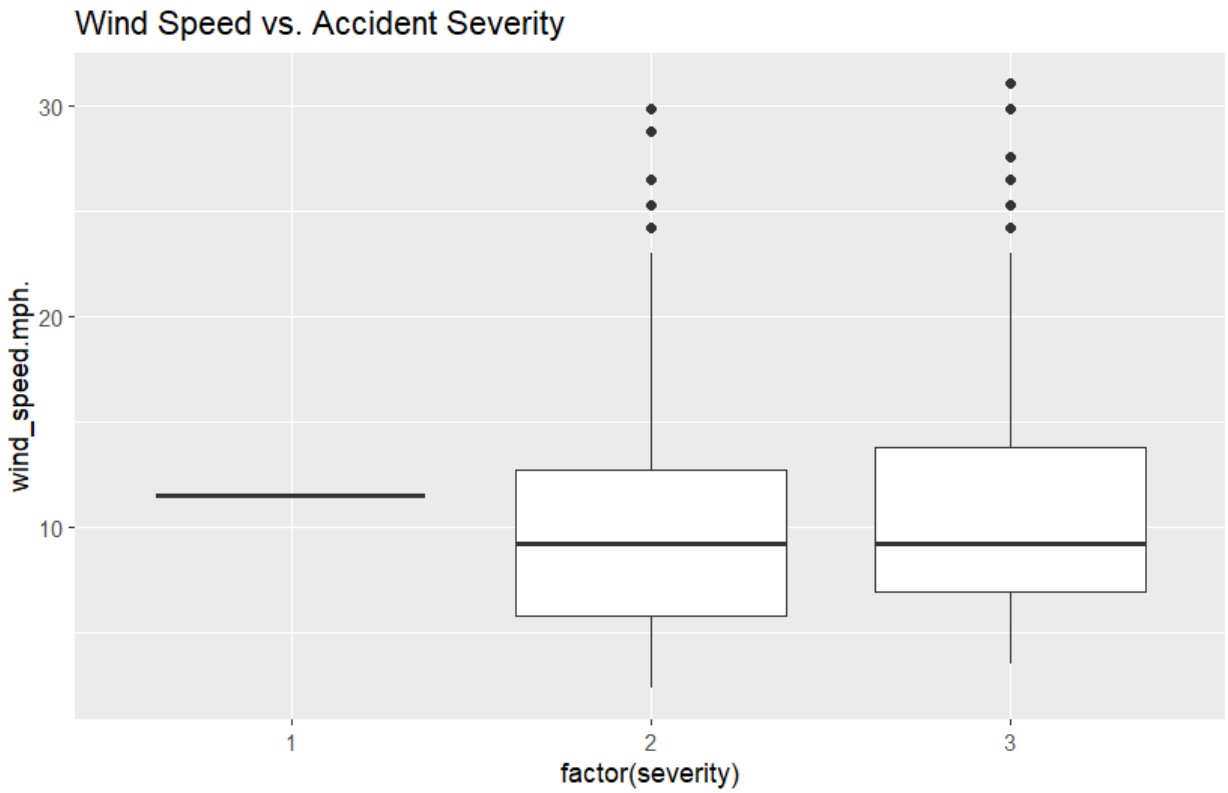
- Evaluate a model to determine the relationship between:
    - Road conditions and frequency
    - Severity of accidents and accident occurrences
- Plots?
    - Boxplots
    - Scatterplots
    - Q-Q plot (normality?) -will be hard to achieve normality, given the nature of the data

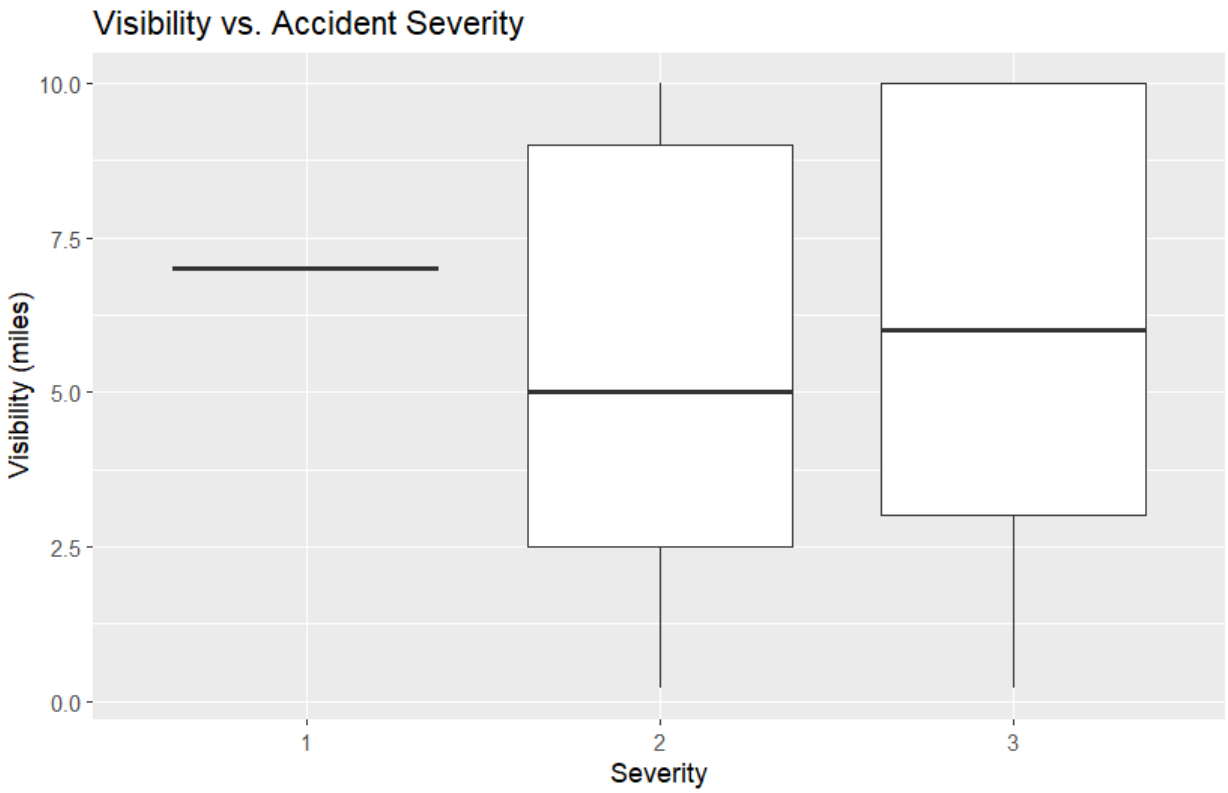
Conclusion

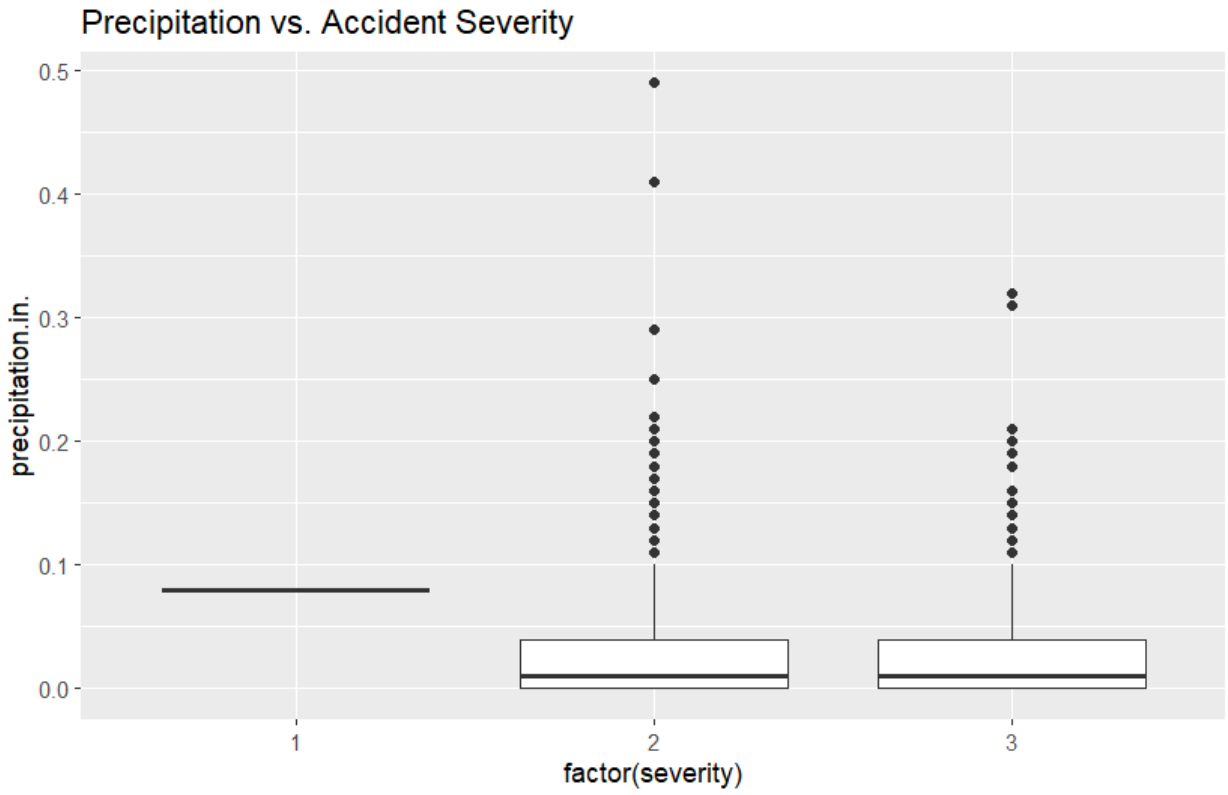
References (for your introduction and data)


Proportion of Accident Severity by Weather Condition
1.00
0.75
0.50
0.25
0.00
Proportion of Severity
Overcast
Rain
Weather Condition
factor(severity)
2
3


Temperature vs. Accident Severity
60
40
20
temperature.f.
1
2
3
factor(severity)


Wind Speed vs. Accident Severity
wind_speed.mph.
30
20
10
1
2
3
factor(severity)






```
data <- read.csv("D:\\US_Accidents_March23_reduced.csv")
```

```
names(data) <- tolower(names(data))

names(data)


# summary of the data

head(data)

str(data)

summary(data)


#removing unneeded columns

data$start_time <- NULL

data$end_time <- NULL

data$end_lat <- NULL

data$end_lng <- NULL

data$wind_chill.f. <- NULL

data$humidity... <- NULL


## making sure columns were removed

names(data)


## what columns have missing data?

colSums(is.na(data))


#how many rows and columns?

nrow(data)

ncol(data)




sum(data$weather_condition == "Light Rain")

sum(data$weather_condition == "Mostly Cloudy")

sum(data$weather_condition == "Overcast")

sum(data$weather_condition == "Light Snow")
```

```
sum(data$weather_condition == "Fog")

sum(data$weather_condition == "Heavy Rain")


mean(data$distance, na.rm = TRUE)


#removing missing #s

df<- data %>%

 filter(!is.na(weather_condition),

       !is.na(precipitation.in.),

       !is.na(visibility.mi.),

       !is.na(wind_speed.mph.),

       !is.na(temperature.f.),

       !is.na(severity))


summary(df)

unique(df$weather_condition)




##relationship between accident severity and weather condition

filtered_data <- df %>% filter(weather_condition %in% c("Overcast", "Rain", "Snow, Clear"))


ggplot(filtered_data, aes(x = weather_condition, fill = factor(severity))) +

 geom_bar(position = "fill") +

 theme(axis.text.x = element_text(angle = 45, hjust = 1)) +

 ggtitle("Proportion of Accident Severity by Weather Condition") +

 xlab("Weather Condition") +

 ylab("Proportion of Severity")


#relationship between temperature and accident severity


ggplot(df, aes(x = factor(severity), y = temperature.f.)) +
```

```
  geom_boxplot() +
  ggtitle("Temperature vs. Accident Severity")


## relationship between wind speed and accident severity
ggplot(df, aes(x = factor(severity), y = wind_speed.mph.)) +
  geom_boxplot() +
  ggtitle("Wind Speed vs. Accident Severity")


##relationship between visibility and accident severity
ggplot(df, aes(x = factor(severity), y = visibility.mi.)) +
  geom_boxplot() +
  ggtitle("Visibility vs. Accident Severity") +
  xlab("Severity") +
  ylab("Visibility (miles)")


## relationship between precipitation and accident severity
ggplot(df, aes(x = factor(severity), y = precipitation.in.)) +
  geom_boxplot() +
  ggtitle("Precipitation vs. Accident Severity")
```