Chicago Traffic Accident Visualization

2024-10-25

most severe injury

```
# read in data
library(readr)
crash_data <- read_csv("crash_data.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 886090 Columns: 48
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (31): CRASH_RECORD_ID, CRASH_DATE_EST_I, CRASH_DATE, TRAFFIC_CONTROL_DEV...
## dbl (17): POSTED_SPEED_LIMIT, LANE_CNT, STREET_NO, BEAT_OF_OCCURRENCE, NUM_U...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Check columns
str(crash_data)
```

```
## spc_tbl_ [886,090 × 48] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ CRASH_RECORD_ID             : chr [1:886090] "6c1659069e9c6285a650e70d6f9b574ed5f64c12
888479093dfeef179c0344ec6d2057eae224b5c0d5dfc278c0a237f8c22543f07fdef2e"| __truncated__ "5f54
a59fcb087b12ae5b1acff96a3caf4f2d37e79f8db4106558b34b8a6d2b81af02cf91b576ecd7ced08ffd10fcfd940
a84f7613125b89"| __truncated__ "61fcb8c1eb522a6469b460e2134df3d15f82e81fd93e9cafd3dc7e631b9e1
ba8b450a63af12bd90d1d2d9b127ea287f88d32e138a4eeba1"| __truncated__ "004cd14d0303a9163aad69a2d
7f341b7da2a8572b2ab3378594bfae8ac53dcb604dd8d414f93c290b55862f9f2517ad32e6209cbc8034c2"| __tr
uncated__ ...
##  $ CRASH_DATE_EST_I            : chr [1:886090] NA NA NA NA ...
##  $ CRASH_DATE                  : chr [1:886090] "08/18/2023 12:50:00 PM" "07/29/2023 02:4
5:00 PM" "08/18/2023 05:58:00 PM" "11/26/2019 08:38:00 AM" ...
##  $ POSTED_SPEED_LIMIT          : num [1:886090] 15 30 30 25 20 30 30 35 30 25 ...
##  $ TRAFFIC_CONTROL_DEVICE      : chr [1:886090] "OTHER" "TRAFFIC SIGNAL" "NO CONTROLS" "N
O CONTROLS" ...
##  $ DEVICE_CONDITION            : chr [1:886090] "FUNCTIONING PROPERLY" "FUNCTIONING PROPE
RLY" "NO CONTROLS" "NO CONTROLS" ...
##  $ WEATHER_CONDITION           : chr [1:886090] "CLEAR" "CLEAR" "CLEAR" "CLEAR" ...
##  $ LIGHTING_CONDITION          : chr [1:886090] "DAYLIGHT" "DAYLIGHT" "DAYLIGHT" "DAYLIGH
T" ...
##  $ FIRST_CRASH_TYPE            : chr [1:886090] "REAR END" "PARKED MOTOR VEHICLE" "PEDALC
YCLIST" "PEDESTRIAN" ...
##  $ TRAFFICWAY_TYPE             : chr [1:886090] "OTHER" "DIVIDED - W/MEDIAN (NOT RAISED)"
"NOT DIVIDED" "ONE-WAY" ...
##  $ LANE_CNT                    : num [1:886090] NA NA NA NA NA NA NA NA NA NA ...
##  $ ALIGNMENT                   : chr [1:886090] "STRAIGHT AND LEVEL" "STRAIGHT AND LEVEL"
"STRAIGHT AND LEVEL" "CURVE ON GRADE" ...
##  $ ROADWAY_SURFACE_COND        : chr [1:886090] "DRY" "DRY" "DRY" "DRY" ...
##  $ ROAD_DEFECT                 : chr [1:886090] "NO DEFECTS" "NO DEFECTS" "NO DEFECTS" "N
O DEFECTS" ...
##  $ REPORT_TYPE                 : chr [1:886090] "ON SCENE" "ON SCENE" "ON SCENE" "ON SCEN
E" ...
##  $ CRASH_TYPE                  : chr [1:886090] "INJURY AND / OR TOW DUE TO CRASH" "NO IN
JURY / DRIVE AWAY" "INJURY AND / OR TOW DUE TO CRASH" "INJURY AND / OR TOW DUE TO CRASH" ...
##  $ INTERSECTION_RELATED_I      : chr [1:886090] NA NA NA NA ...
##  $ NOT_RIGHT_OF_WAY_I          : chr [1:886090] NA NA NA NA ...
##  $ HIT_AND_RUN_I               : chr [1:886090] NA "Y" NA NA ...
##  $ DAMAGE                      : chr [1:886090] "OVER $1,500" "OVER $1,500" "$501 - $1,50
0" "OVER $1,500" ...
##  $ DATE_POLICE_NOTIFIED        : chr [1:886090] "08/18/2023 12:55:00 PM" "07/29/2023 02:4
5:00 PM" "08/18/2023 06:01:00 PM" "11/26/2019 08:38:00 AM" ...
##  $ PRIM_CONTRIBUTORY_CAUSE     : chr [1:886090] "FOLLOWING TOO CLOSELY" "FAILING TO REDUC
E SPEED TO AVOID CRASH" "FAILING TO REDUCE SPEED TO AVOID CRASH" "UNABLE TO DETERMINE" ...
##  $ SEC_CONTRIBUTORY_CAUSE      : chr [1:886090] "DISTRACTION - FROM INSIDE VEHICLE" "OPER
ATING VEHICLE IN ERRATIC, RECKLESS, CARELESS, NEGLIGENT OR AGGRESSIVE MANNER" "UNABLE TO DETE
RMINE" "NOT APPLICABLE" ...
##  $ STREET_NO                   : num [1:886090] 700 2101 3422 5 3 ...
##  $ STREET_DIRECTION            : chr [1:886090] "W" "S" "N" "W" ...
##  $ STREET_NAME                 : chr [1:886090] "OHARE ST" "ASHLAND AVE" "LONG AVE" "TERM
INAL ST" ...
##  $ BEAT_OF_OCCURRENCE          : num [1:886090] 1654 1235 1633 1655 1653 ...
##  $ PHOTOS_TAKEN_I              : chr [1:886090] NA NA NA "Y" ...
##  $ STATEMENTS_TAKEN_I          : chr [1:886090] NA NA NA "Y" ...
##  $ DOORING_I                   : chr [1:886090] NA NA NA NA ...
##  $ WORK_ZONE_I                 : chr [1:886090] NA NA NA NA ...
##  $ WORK_ZONE_TYPE              : chr [1:886090] NA NA NA NA ...
```

```
##  $ WORKERS_PRESENT_I            : chr [1:886090] NA NA NA NA ...
##  $ NUM_UNITS                    : num [1:886090] 2 4 2 2 1 2 2 2 2 2 ...
##  $ MOST_SEVERE_INJURY           : chr [1:886090] "NONINCAPACITATING INJURY" "NO INDICATION
OF INJURY" "NONINCAPACITATING INJURY" "FATAL" ...
##  $ INJURIES_TOTAL               : num [1:886090] 1 0 1 1 0 0 0 0 0 0 ...
##  $ INJURIES_FATAL               : num [1:886090] 0 0 0 1 0 0 0 0 0 0 ...
##  $ INJURIES_INCAPACITATING      : num [1:886090] 0 0 0 0 0 0 0 0 0 0 ...
##  $ INJURIES_NON_INCAPACITATING  : num [1:886090] 1 0 1 0 0 0 0 0 0 0 ...
##  $ INJURIES_REPORTED_NOT_EVIDENT: num [1:886090] 0 0 0 0 0 0 0 0 0 0 ...
##  $ INJURIES_NO_INDICATION       : num [1:886090] 1 1 1 1 1 3 2 2 2 5 ...
##  $ INJURIES_UNKNOWN             : num [1:886090] 0 0 0 0 0 0 0 0 0 0 ...
##  $ CRASH_HOUR                   : num [1:886090] 12 14 17 8 10 13 17 13 0 19 ...
##  $ CRASH_DAY_OF_WEEK            : num [1:886090] 6 7 6 3 6 7 2 1 1 2 ...
##  $ CRASH_MONTH                  : num [1:886090] 8 7 8 11 8 7 2 8 8 1 ...
##  $ LATITUDE                     : num [1:886090] NA 41.9 41.9 NA NA ...
##  $ LONGITUDE                    : num [1:886090] NA -87.7 -87.8 NA NA ...
##  $ LOCATION                     : chr [1:886090] NA "POINT (-87.665902342962 41.8541202629
52)" "POINT (-87.761883496974 41.942975745006)" NA ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   CRASH_RECORD_ID = col_character(),
##   ..   CRASH_DATE_EST_I = col_character(),
##   ..   CRASH_DATE = col_character(),
##   ..   POSTED_SPEED_LIMIT = col_double(),
##   ..   TRAFFIC_CONTROL_DEVICE = col_character(),
##   ..   DEVICE_CONDITION = col_character(),
##   ..   WEATHER_CONDITION = col_character(),
##   ..   LIGHTING_CONDITION = col_character(),
##   ..   FIRST_CRASH_TYPE = col_character(),
##   ..   TRAFFICWAY_TYPE = col_character(),
##   ..   LANE_CNT = col_double(),
##   ..   ALIGNMENT = col_character(),
##   ..   ROADWAY_SURFACE_COND = col_character(),
##   ..   ROAD_DEFECT = col_character(),
##   ..   REPORT_TYPE = col_character(),
##   ..   CRASH_TYPE = col_character(),
##   ..   INTERSECTION_RELATED_I = col_character(),
##   ..   NOT_RIGHT_OF_WAY_I = col_character(),
##   ..   HIT_AND_RUN_I = col_character(),
##   ..   DAMAGE = col_character(),
##   ..   DATE_POLICE_NOTIFIED = col_character(),
##   ..   PRIM_CONTRIBUTORY_CAUSE = col_character(),
##   ..   SEC_CONTRIBUTORY_CAUSE = col_character(),
##   ..   STREET_NO = col_double(),
##   ..   STREET_DIRECTION = col_character(),
##   ..   STREET_NAME = col_character(),
##   ..   BEAT_OF_OCCURRENCE = col_double(),
##   ..   PHOTOS_TAKEN_I = col_character(),
##   ..   STATEMENTS_TAKEN_I = col_character(),
##   ..   DOORING_I = col_character(),
##   ..   WORK_ZONE_I = col_character(),
##   ..   WORK_ZONE_TYPE = col_character(),
##   ..   WORKERS_PRESENT_I = col_character(),
##   ..   NUM_UNITS = col_double(),
##   ..   MOST_SEVERE_INJURY = col_character(),
##   ..   INJURIES_TOTAL = col_double(),
```

```
##   ..   INJURIES_FATAL = col_double(),
##   ..   INJURIES_INCAPACITATING = col_double(),
##   ..   INJURIES_NON_INCAPACITATING = col_double(),
##   ..   INJURIES_REPORTED_NOT_EVIDENT = col_double(),
##   ..   INJURIES_NO_INDICATION = col_double(),
##   ..   INJURIES_UNKNOWN = col_double(),
##   ..   CRASH_HOUR = col_double(),
##   ..   CRASH_DAY_OF_WEEK = col_double(),
##   ..   CRASH_MONTH = col_double(),
##   ..   LATITUDE = col_double(),
##   ..   LONGITUDE = col_double(),
##   ..   LOCATION = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
library(ggplot2)
```

```
## Warning: 程辑包'ggplot2'是用R版本4.3.3 来建造的
```

```
library(tidyverse)
```

```
## Warning: 程辑包'purrr'是用R版本4.3.3 来建造的
```

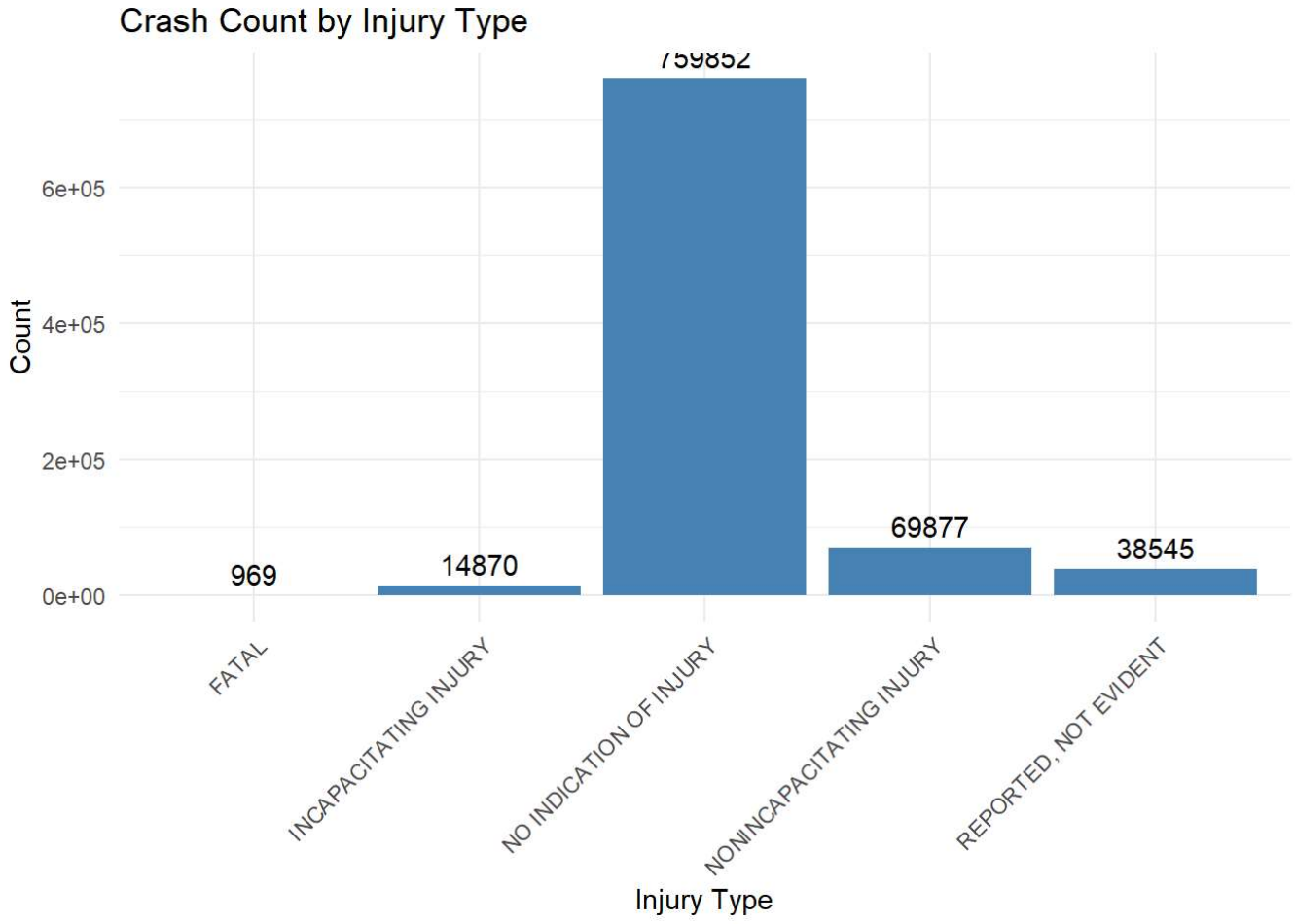
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.2     ✓ stringr   1.5.0
## ✓ forcats   1.0.0     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```
# Injury type: check unique values and filter them out
unique(crash_data$MOST_SEVERE_INJURY)
```

```
## [1] "NONINCAPACITATING INJURY" "NO INDICATION OF INJURY"
## [3] "FATAL"                    "REPORTED, NOT EVIDENT"
## [5] "INCAPACITATING INJURY"    NA
## [7] "2"                        "0"
## [9] "-87.81704561"
```

```
filtered_data1 <- crash_data %>% filter(!MOST_SEVERE_INJURY %in% c("2", "0","-87.81704561",N
A))

# bar plot: injury
ggplot(filtered_data1, aes(x = factor(MOST_SEVERE_INJURY))) +
  geom_bar(fill = "steelblue") +
  geom_text(stat = 'count', aes(label = ..count..), vjust = -0.5) +
  labs(title = "Crash Count by Injury Type",
       x = "Injury Type",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

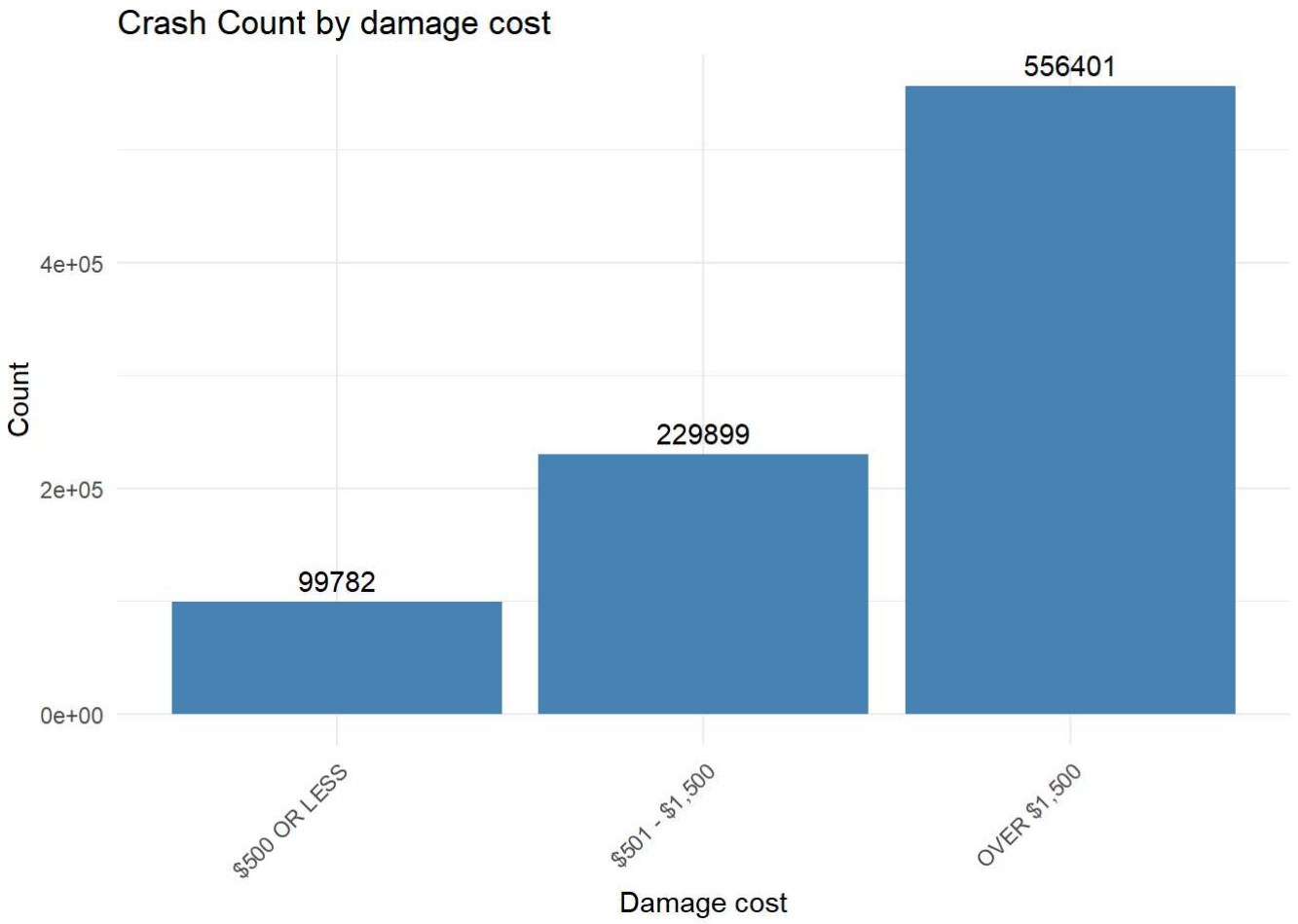
```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## ℹ Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
# damage:~
unique(crash_data$DAMAGE)
```

```
## [1] "OVER $1,500"
## [2] "$501 - $1,500"
## [3] "$500 OR LESS"
## [4] "O\b\xb0U\x95\x94\031\x90\x89:wFz\xdf\xc3\006\x98S\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\x
ccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3
|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt
\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027
\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\x
e3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbb
t\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\02
7\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ
\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\x
bbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020
\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xf
fJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67
\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\02
0\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\x
ffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe6
7\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\0
20\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS
\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\x
e67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#
\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccO
S\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F
\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe
5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xc
cOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|
F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\x
e5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\x
ccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3
|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt
\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027
\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\x
e3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbb
t\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\02
7\xccOS\xffJ\xe3|F\xe67\xbbt\xe5#\020\027\xccOh\xad\0
33V\xc9\xc8\xfe\x8e\xb0\034\x9b\x88Q\x8f\xb8\xd287.629386933,POINT (-87.629386933044 41.85299
1219253)\r\nde6fa485bd3ad647f3524d1c615f65692eefa4b06487887f4cab60ed19e6e1e9611197eb4b3dce382
7bc87a60455d827aafecdcf18cf3c7e6700b1da74058907,,04/17/2022 02:20:00 PM,35,TRAFFIC SIGNAL,FUN
CTIONING PROPERLY,CLEAR,DAYLIGHT,REAR END,NOT DIVIDED,,STRAIGHT AND LEVEL,DRY,NO DEFECTS,,INJ
URY AND / OR TOW DUE TO CRASH,,,,OVER $1"
## [5] NA
## [6] "331"
## [7] "NO INDICATION OF INJURY"
## [8] "1"
```

```
filtered_data2 <- crash_data %>% filter(DAMAGE %in% c("OVER $1,500", "$501 - $1,500","$500 OR
LESS"))

# Bar plot: damage
ggplot(filtered_data2, aes(x = DAMAGE)) +
  geom_bar(fill = "steelblue") +
  geom_text(stat = 'count', aes(label = ..count..), vjust = -0.5) +
  labs(title = "Crash Count by damage cost",
       x = "Damage cost",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



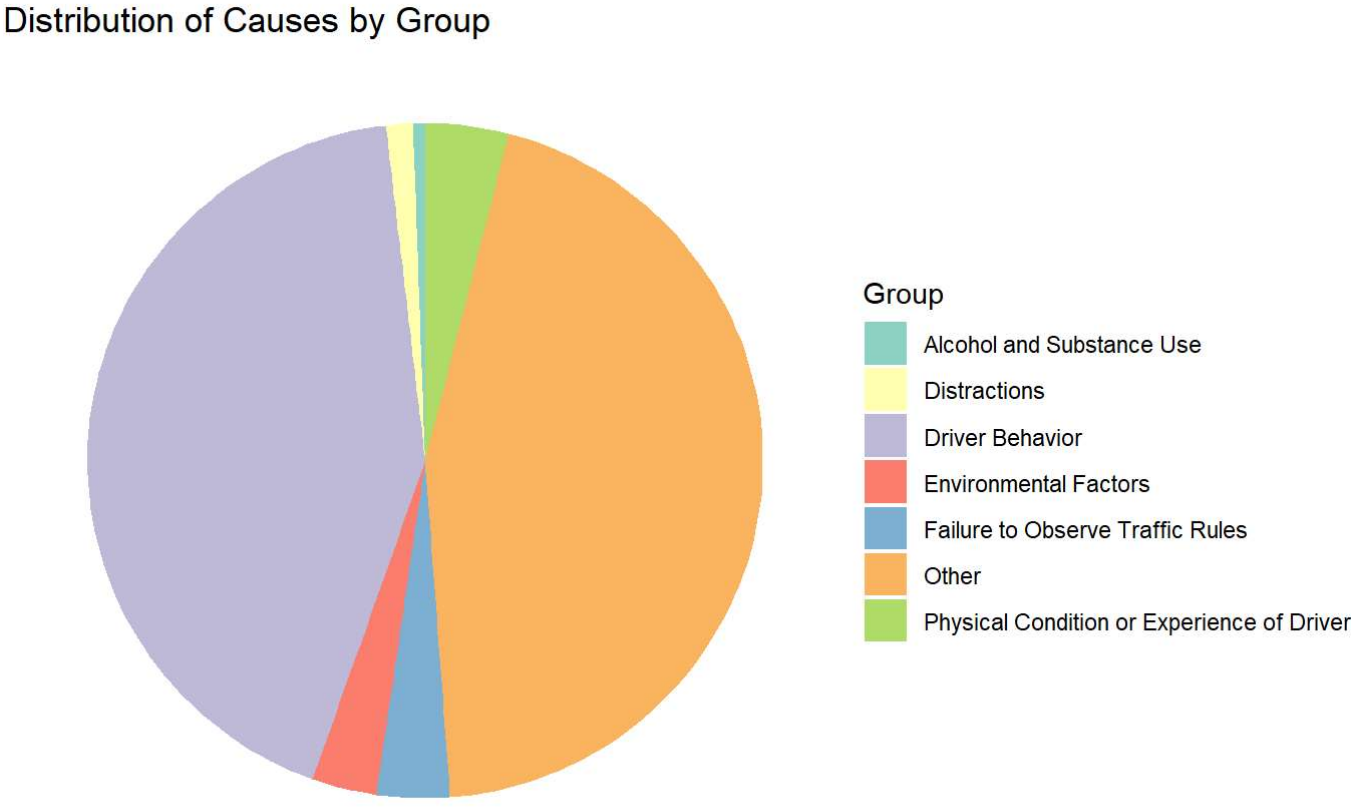

```
# prim_contributory: ~; exclude abnormal values
unique(crash_data$PRIM_CONTRIBUTORY_CAUSE)
```

```
##  [1] "FOLLOWING TOO CLOSELY"
##  [2] "FAILING TO REDUCE SPEED TO AVOID CRASH"
##  [3] "UNABLE TO DETERMINE"
##  [4] "IMPROPER BACKING"
##  [5] "IMPROPER TURNING/NO SIGNAL"
##  [6] "NOT APPLICABLE"
##  [7] "WEATHER"
##  [8] "IMPROPER OVERTAKING/PASSING"
##  [9] "DRIVING SKILLS/KNOWLEDGE/EXPERIENCE"
## [10] "IMPROPER LANE USAGE"
## [11] "VISION OBSCURED (SIGNS, TREE LIMBS, BUILDINGS, ETC.)"
## [12] "ROAD ENGINEERING/SURFACE/MARKING DEFECTS"
## [13] "FAILING TO YIELD RIGHT-OF-WAY"
## [14] "EQUIPMENT - VEHICLE CONDITION"
## [15] "RELATED TO BUS STOP"
## [16] "DISREGARDING OTHER TRAFFIC SIGNS"
## [17] "DRIVING ON WRONG SIDE/WRONG WAY"
## [18] "ROAD CONSTRUCTION/MAINTENANCE"
## [19] "DISTRACTION - FROM INSIDE VEHICLE"
## [20] "ANIMAL"
## [21] "TEXTING"
## [22] "DISREGARDING TRAFFIC SIGNALS"
## [23] "DISREGARDING ROAD MARKINGS"
## [24] "CELL PHONE USE OTHER THAN TEXTING"
## [25] "DISREGARDING STOP SIGN"
## [26] "OPERATING VEHICLE IN ERRATIC, RECKLESS, CARELESS, NEGLIGENT OR AGGRESSIVE MANNER"
## [27] "EXCEEDING AUTHORIZED SPEED LIMIT"
## [28] "DISTRACTION - FROM OUTSIDE VEHICLE"
## [29] "PHYSICAL CONDITION OF DRIVER"
## [30] "EXCEEDING SAFE SPEED FOR CONDITIONS"
## [31] "DISREGARDING YIELD SIGN"
## [32] "TURNING RIGHT ON RED"
## [33] "UNDER THE INFLUENCE OF ALCOHOL/DRUGS (USE WHEN ARREST IS EFFECTED)"
## [34] "EVASIVE ACTION DUE TO ANIMAL, OBJECT, NONMOTORIST"
## [35] "HAD BEEN DRINKING (USE WHEN ARREST IS NOT MADE)"
## [36] "DISTRACTION - OTHER ELECTRONIC DEVICE (NAVIGATION DEVICE, DVD PLAYER, ETC.)"
## [37] "OBSTRUCTED CROSSWALKS"
## [38] "BICYCLE ADVANCING LEGALLY ON RED LIGHT"
## [39] "PASSING STOPPED SCHOOL BUS"
## [40] "MOTORCYCLE ADVANCING LEGALLY ON RED LIGHT"
## [41] "04/17/2022 02:23:00 PM"
## [42] NA
## [43] "0"
## [44] "2"
```

```
filtered_data3 <- crash_data %>% filter(!PRIM_CONTRIBUTORY_CAUSE %in% c("04/17/2022 02:23:00
PM",NA,"0","2"))

# Define the groups based on the prim_cause column
filtered_data3 <- filtered_data3 %>%
  mutate(Group = case_when(
    # Driver Behavior
    PRIM_CONTRIBUTORY_CAUSE %in% c("FOLLOWING TOO CLOSELY", "FAILING TO REDUCE SPEED TO AVOID
CRASH",
                "IMPROPER TURNING/NO SIGNAL", "IMPROPER OVERTAKING/PASSING",
                "FAILING TO YIELD RIGHT-OF-WAY", "IMPROPER LANE USAGE",
                "EXCEEDING AUTHORIZED SPEED LIMIT",
                "OPERATING VEHICLE IN ERRATIC, RECKLESS, CARELESS, NEGLIGENT OR AGGRESSIVE M
ANNER",
                "IMPROPER BACKING",
                "DRIVING ON WRONG SIDE/WRONG WAY",
                "RELATED TO BUS STOP") ~ "Driver Behavior",

    # Distractions
    PRIM_CONTRIBUTORY_CAUSE %in% c("DISTRACTION - FROM INSIDE VEHICLE", "DISTRACTION - FROM O
UTSIDE VEHICLE",
                "DISTRACTION - OTHER ELECTRONIC DEVICE (NAVIGATION DEVICE, DVD PLAYER, ET
C.)",
                "TEXTING", "CELL PHONE USE OTHER THAN TEXTING") ~ "Distractions",

    # Environmental Factors
    PRIM_CONTRIBUTORY_CAUSE %in% c("WEATHER", "VISION OBSCURED (SIGNS, TREE LIMBS, BUILDINGS,
ETC.)",
                "ROAD ENGINEERING/SURFACE/MARKING DEFECTS", "ROAD CONSTRUCTION/MAINTENANCE",
                "EQUIPMENT - VEHICLE CONDITION",
                "OBSTRUCTED CROSSWALKS") ~ "Environmental Factors",

    # Failure to Observe Traffic Rules
    PRIM_CONTRIBUTORY_CAUSE %in% c("DISREGARDING TRAFFIC SIGNALS", "DISREGARDING STOP SIGN",
                "DISREGARDING YIELD SIGN", "DISREGARDING ROAD MARKINGS","DISREGARDING OTHER
TRAFFIC SIGNS",
                "TURNING RIGHT ON RED", "TURNING LEFT ON RED",
                "MOTORCYCLE ADVANCING LEGALLY ON RED LIGHT",
                "BICYCLE ADVANCING LEGALLY ON RED LIGHT",
                "PASSING STOPPED SCHOOL BUS") ~ "Failure to Observe Traffic Rules",

    # Alcohol and Substance Use
    PRIM_CONTRIBUTORY_CAUSE %in% c("UNDER THE INFLUENCE OF ALCOHOL/DRUGS (USE WHEN ARREST IS
EFFECTED)",
                "HAD BEEN DRINKING (USE WHEN ARREST IS NOT MADE)") ~ "Alcohol and Substance
Use",

    # Physical Condition or Experience of Driver
    PRIM_CONTRIBUTORY_CAUSE %in% c("DRIVING SKILLS/KNOWLEDGE/EXPERIENCE",
                "PHYSICAL CONDITION OF DRIVER") ~ "Physical Condition or Experience of Drive
r",

    # other
    PRIM_CONTRIBUTORY_CAUSE %in% c("UNABLE TO DETERMINE", "NOT APPLICABLE",
                "EVASIVE ACTION DUE TO ANIMAL, OBJECT, NONMOTORIST",
```

```
                "ANIMAL", "EXCEEDING SAFE SPEED FOR CONDITIONS"
                ) ~ "Other"

  ))
```

```
# count by group
reason_count <- filtered_data3 %>%
  count(Group) %>%
  mutate(percentage = round((n/sum(n)*100),3))

# general pie chart
ggplot(reason_count, aes(x = "", y = n, fill = Group)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Distribution of Causes by Group") +
  scale_fill_brewer(palette = "Set3") +
  theme_void() +
  theme(legend.position = "right")
```



```
# since Other part takes a large section, building another pie chart to dive-in
# filter and count
other_data <- subset(filtered_data3, Group == "Other")
other_reason_count <- other_data %>%
  count(PRIM_CONTRIBUTORY_CAUSE) %>%
  mutate(percentage = round((n/sum(n)*100),3))

# find out that "useless" values took the most percentage of it
# build a dive-in pie chart without NOT APPLICABLE and UNABLE TO DETERMINE
clean_other_reason_count <- other_reason_count %>%
  filter(!PRIM_CONTRIBUTORY_CAUSE %in% c("NOT APPLICABLE","UNABLE TO DETERMINE"))

ggplot(clean_other_reason_count, aes(x = "", y = n, fill = PRIM_CONTRIBUTORY_CAUSE)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  ggtitle("Distribution of Causes within 'Other' Group")
```