yanyue-02-analysis

YanyueFu (U41816127)
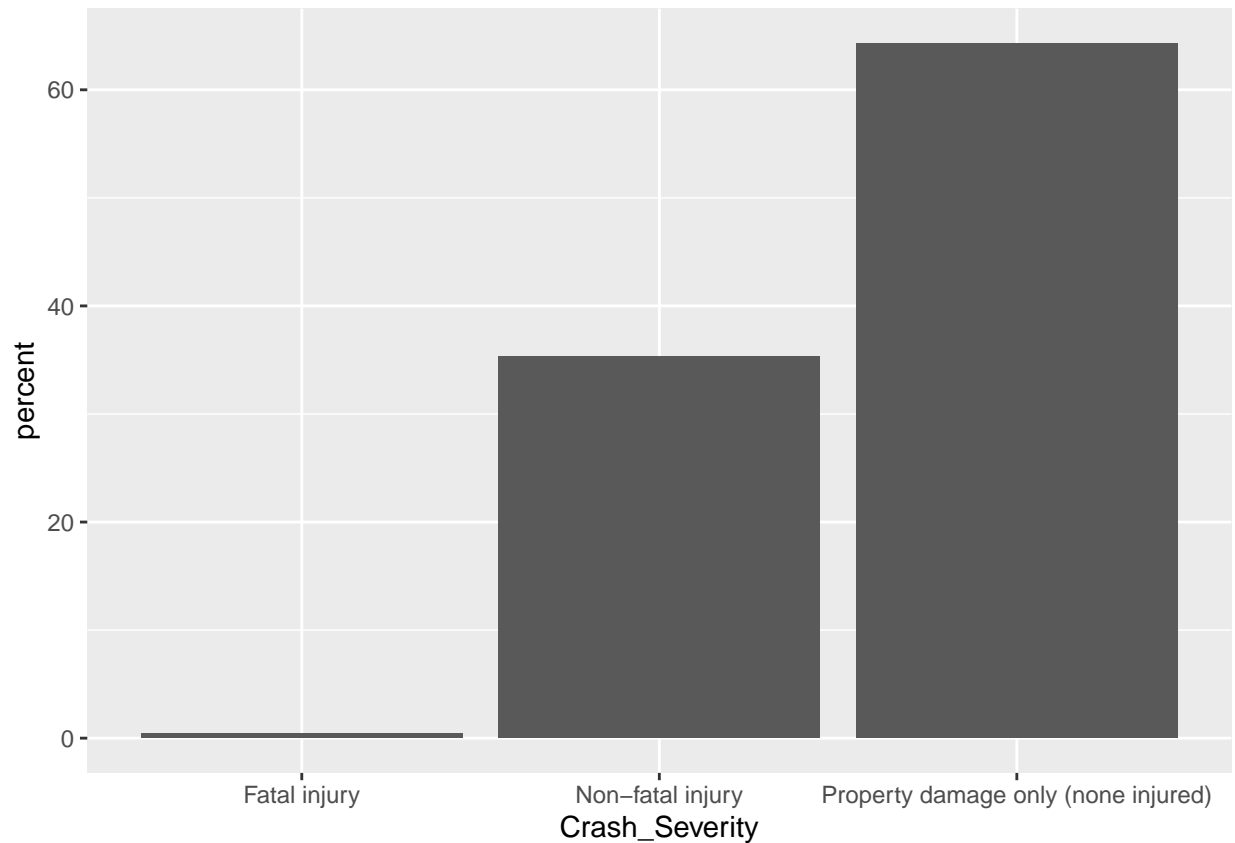
3/18/2020

####clean the data

```r
data = fread("crash-updated.csv")
data_clean = data %>% filter(!Crash_Severity %in% c('Unknown',
                                                     'Not Reported'),
                             !Maximum_Injury_Severity_Reported %in%
                               c('Unknown',
                                 'Not Reported'),
                             !Manner_of_Collision %in%
                               c('Unknown','Not Reported',"",
                                 'Not reported'))
write.csv(data_clean, "cleaned.csv")
```

```r
collision = data_clean  %>%  select(Crash_Severity) %>%
  group_by(Crash_Severity) %>%
  count() %>%
  mutate(percent = round(freq / sum(freq) * 100,1)) %>%
  mutate(Crash_Severity = reorder(Crash_Severity,percent))

collision %>% ggplot(aes(Crash_Severity,percent))+
  geom_bar(stat = 'identity')
```
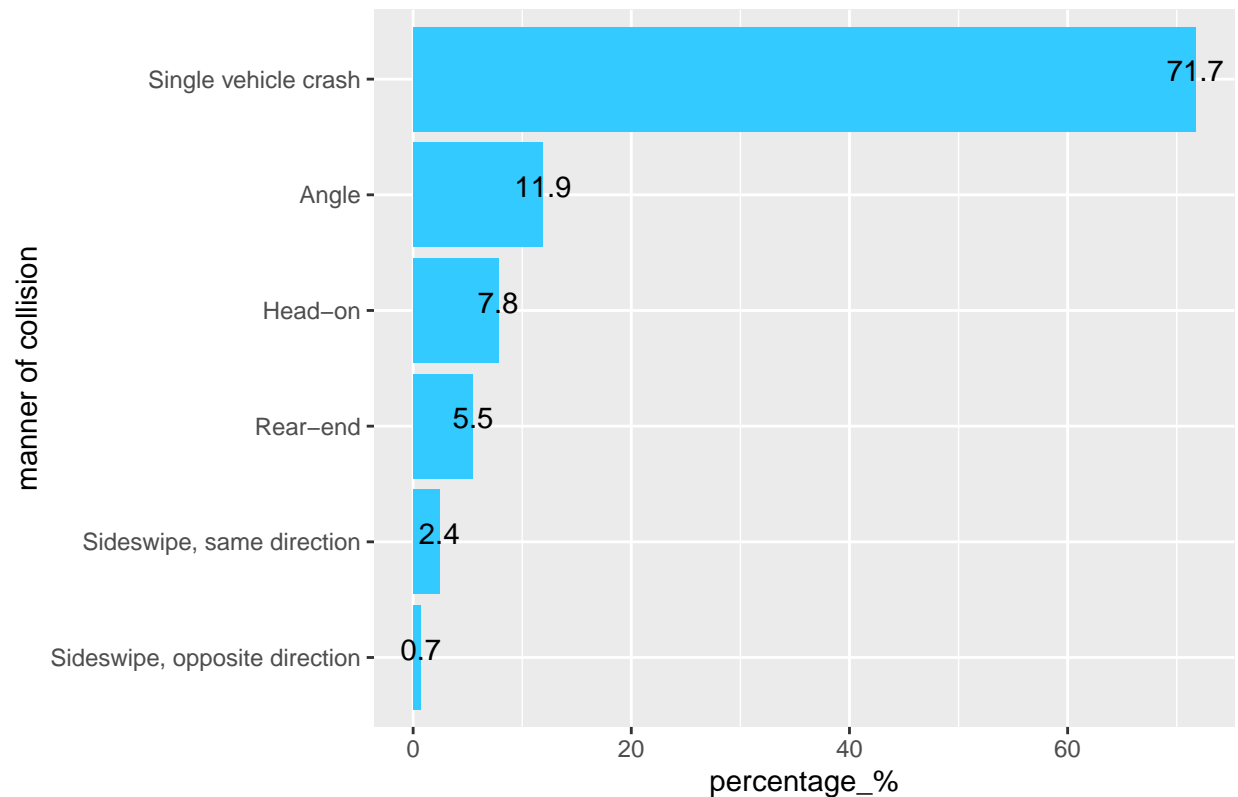
```r
# Hist of each manner of collision caused fatal injury
fatal = data_clean %>% filter(Crash_Severity == "Fatal injury") %>%
  select(Manner_of_Collision) %>%
  group_by(Manner_of_Collision) %>%
  count() %>%
  mutate(percent = round(freq / sum(freq) * 100,1)) %>%
  mutate(Manner_of_Collision = reorder(Manner_of_Collision,percent))

fatal %>% ggplot(aes(Manner_of_Collision,percent))+
  geom_bar(stat = 'identity',fill = '#33CAFF')+
  coord_flip()+
  labs(x = 'manner of collision',
       y = 'percentage_%',
       title = 'The percentage of each collision manner causes fatal injury') +geom_text(aes(label = pe:
```
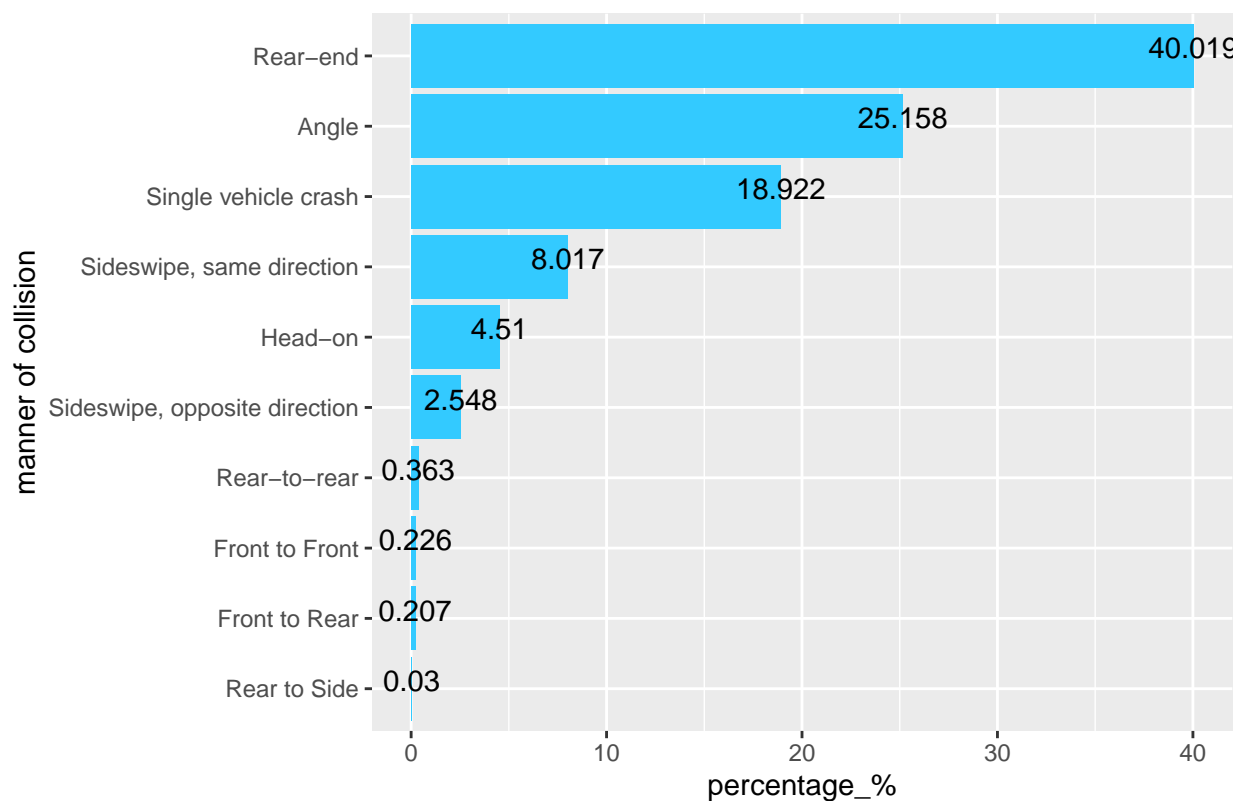
The percentage of each collision manner causes fatal inju


```
nonfatal = data_clean %>% filter(Crash_Severity == "Non-fatal injury") %>%
  select(Manner_of_Collision) %>%
  group_by(Manner_of_Collision) %>%
  count() %>%
  mutate(percent = round(freq / sum(freq) * 100, 3)) %>%
  mutate(Manner_of_Collision = reorder(Manner_of_Collision,percent))

nonfatal %>% ggplot(aes(Manner_of_Collision,percent))+
  geom_bar(stat = 'identity',fill = '#33CAFF')+
  coord_flip()+
  labs(x = 'manner of collision',
       y = 'percentage_%',
       title = 'The percentage of each collision manner causes non-fatal injury') +geom_text(aes(label =
```
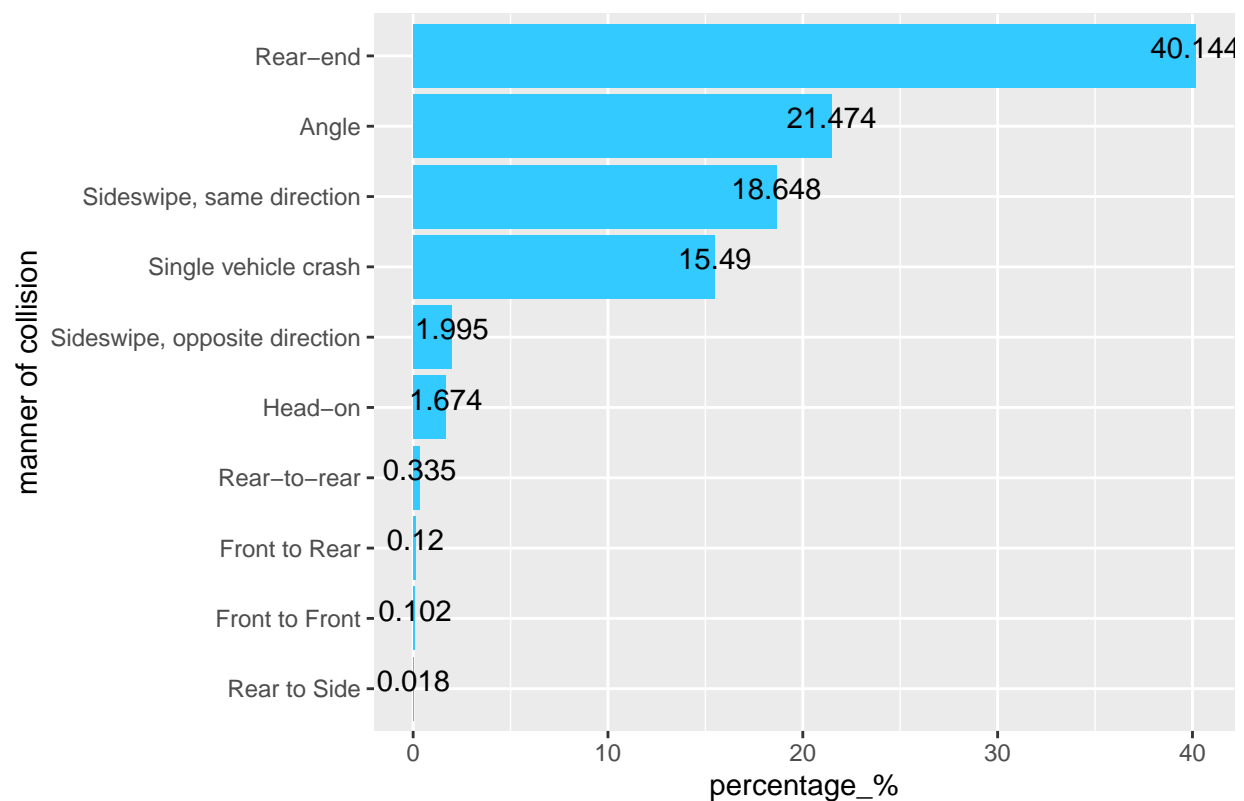
The percentage of each collision manner causes non-fat[...]


```
property = data_clean %>% filter(Crash_Severity == "Property damage only (none injured)") %>%
  select(Manner_of_Collision) %>%
  group_by(Manner_of_Collision) %>%
  count() %>%
  mutate(percent = round(freq / sum(freq) * 100, 3)) %>%
  mutate(Manner_of_Collision = reorder(Manner_of_Collision,percent))

property %>% ggplot(aes(Manner_of_Collision,percent))+
  geom_bar(stat = 'identity',fill = '#33CAFF')+
  coord_flip()+
  labs(x = 'manner of collision',
       y = 'percentage_%',
       title = 'The percentage of collision manner causes Property damage') +geom_text(aes(label = perc[...]
```
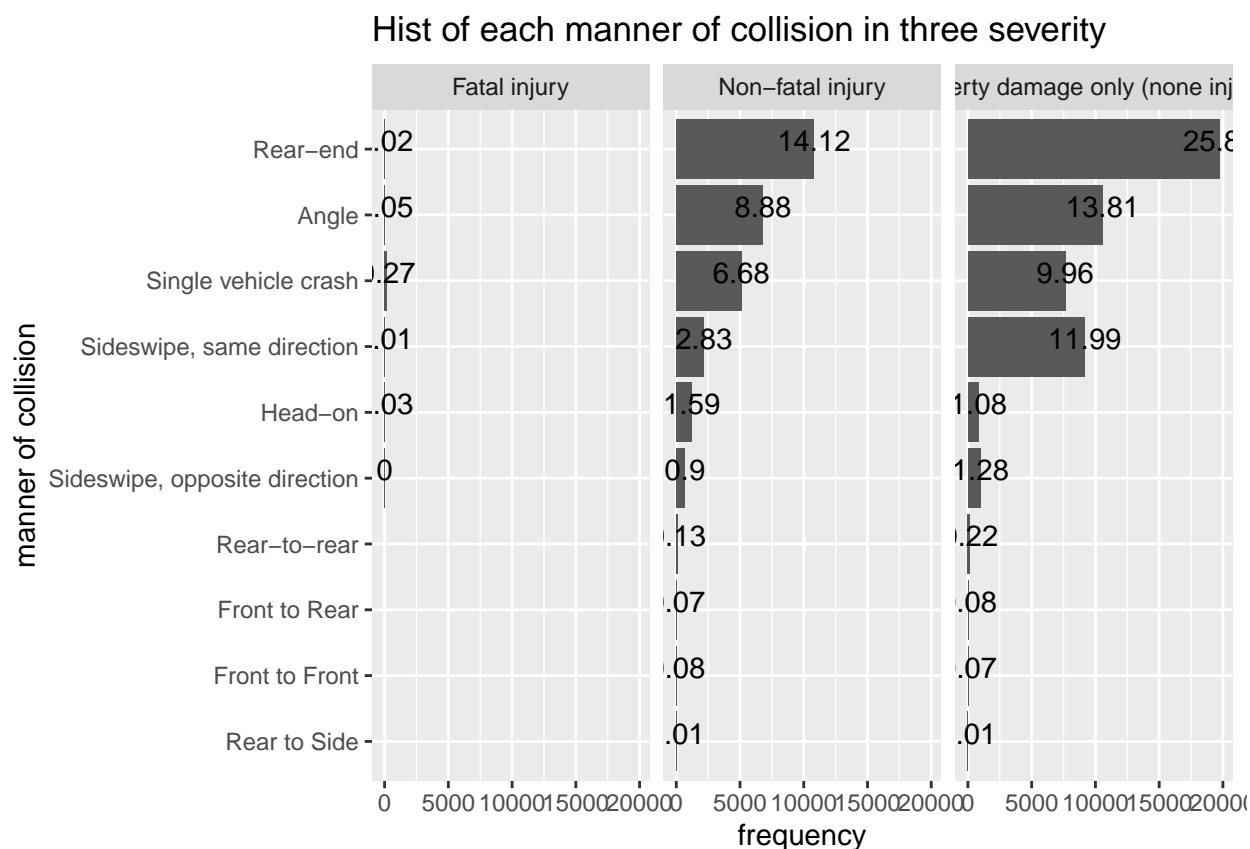
The percentage of collision manner causes Property dam


```r
# hist for the manner of collision
collision = data_clean %>% select(Manner_of_Collision,Crash_Severity) %>%
  group_by(Manner_of_Collision,Crash_Severity) %>%
  count() %>%
  mutate(percent = round(freq / sum(freq) * 100,2)) %>%
  mutate(Manner_of_Collision = reorder(Manner_of_Collision,freq)) %>% arrange(desc(freq))

collision %>% ggplot(aes(Manner_of_Collision,freq))+
  geom_bar(stat = 'identity')+
  facet_grid(~Crash_Severity) +
  coord_flip()+
  labs(x = 'manner of collision',
       y = 'frequency',
       title = 'Hist of each manner of collision in three severity') +
  geom_text(aes(label = percent),vjust = 0.1)
```

Hist of each manner of collision in three severity

	Fatal injury	Non–fatal injury	erty damage only (none inj
Rear–end	.02	14.12	25.8
Angle	.05	8.88	13.81
Single vehicle crash	.27	6.68	9.96
Sideswipe, same direction	.01	2.83	11.99
Head–on	.03	1.59	1.08
Sideswipe, opposite direction	0	0.9	1.28
Rear–to–rear		.13	.22
Front to Rear		.07	.08
Front to Front		.08	.07
Rear to Side		.01	.01

About the collision, the first is fata injury based on different manners of collision. As we can see from the graph, most of the people dead because of single vehicle crash, while angle crash is also serious (except not reported one).

And next is non-fata injury based on different manners of collision. From the bar chart, most people got hurt as a result of rear-end crash while angle crash still stood on the second. Moreover, the severity of single-vehicle-crash also cannot be ignored.

Most vehicle crash are property damage only and the top three collision manners are read-end, angle and sideswipe (with same direction).In summary, no matter people died, got injury or just property damage only in a car accident, rear-end crash and angle crash is something we can't underestimate.

Here are the top ten type of cars involved in the crash, it shows that the almost 95 percent of the vehicle crash caused by the passenger car and the light trucks like van. Notably, the passenger cars has the highest percentage-66 percent, which we need pay more attention to.