mass

SaiNagaChandraVivekGarimella

August 16, 2019

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------- tidyverse 1
```

```
## v ggplot2 3.1.0     v readr   1.3.1
## v tibble  1.4.2     v purrr   0.2.5
## v tidyr   0.8.2     v stringr 1.3.1
## v ggplot2 3.1.0     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------------------------------- tidyverse_conflic
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
main<-read_csv('C:/Users/garim/OneDrive/Desktop/NEU/Introduction to Data Management and Processing/data
```

```
## Parsed with column specification:
## cols(
##   LONCOD = col_double(),
##   LATCOD = col_double(),
##   CDCODE = col_double(),
##   year = col_double(),
##   school = col_character(),
##   city = col_character(),
##   state = col_character(),
##   address = col_character(),
##   urbanrural = col_character(),
##   race = col_character(),
##   killed = col_double(),
##   injured = col_double(),
```

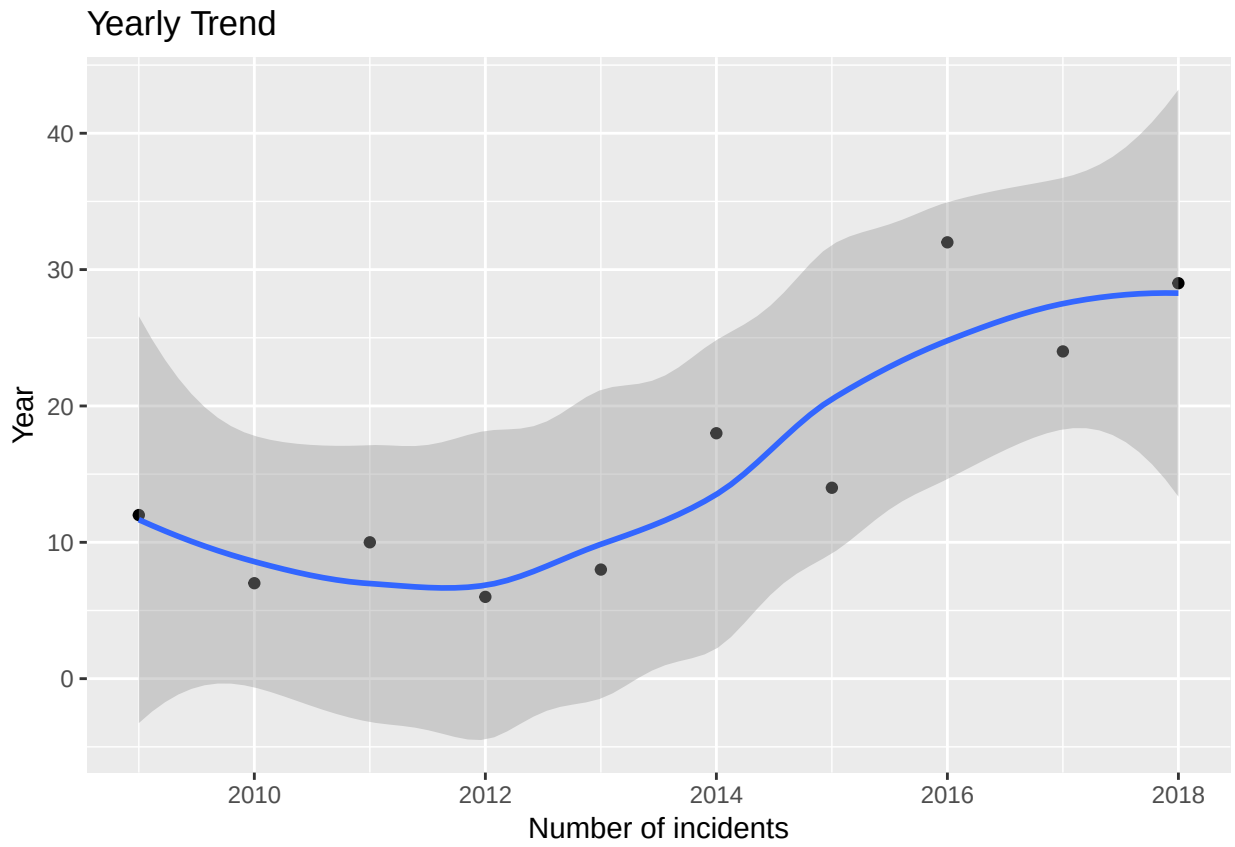
```
##   victims = col_double(),
##   date = col_character(),
##   type = col_character(),
##   NCESSCH = col_double(),
##   time = col_time(format = "")
## )
```

```
main<-drop_na(main)
main
```

```
## # A tibble: 160 x 17
##    LONCOD LATCOD CDCODE  year school city  state address urbanrural race
##     <dbl>  <dbl>  <dbl> <dbl> <chr>  <chr> <chr> <chr>   <chr>      <chr>
##  1  -75.6   39.7   1000  2009 Willi~ New ~ Dela~ 713 E ~ suburbtown BLACK
##  2  -87.6   41.8   1701  2009 Dunba~ Chic~ Illi~ 3000 S~ urban      BLACK
##  3  -90.2   38.6   1712  2009 Cahok~ Caho~ Illi~ 800 Ra~ suburbtown BLACK
##  4  -78.3   35.8   3713  2009 Zebul~ Zebu~ Nort~ Halifa~ suburbtown BLACK
##  5  -79.0   35.1   3702  2009 Westo~ Faye~ Nort~ 277 Bo~ urban      BLACK
##  6  -77.4   37.5   5103  2009 Chimb~ Rich~ Virg~ 3000 E~ urban      BLACK
##  7  -84.5   38.1   2106  2009 Leest~ Lexi~ Kent~ 2010 L~ urban      BLACK
##  8  -92.9   42.6   1901  2009 Aplin~ Park~ Iowa  610 N ~ rural      WHITE
##  9  -72.5   41.0   3601  2009 Matti~ Matt~ New ~ 15125 ~ suburbtown WHITE
## 10 -118.    33.8    637  2009 Wilso~ Long~ Cali~ 4400 E~ urban      HISP
## # ... with 150 more rows, and 7 more variables: killed <dbl>,
## #   injured <dbl>, victims <dbl>, date <chr>, type <chr>, NCESSCH <dbl>,
## #   time <time>
```

```
#YEAR SMOOTH
main %>% group_by(year) %>% summarise(total=n()) %>% ggplot(mapping = aes(x=year,y=total))+geom_point()+
```

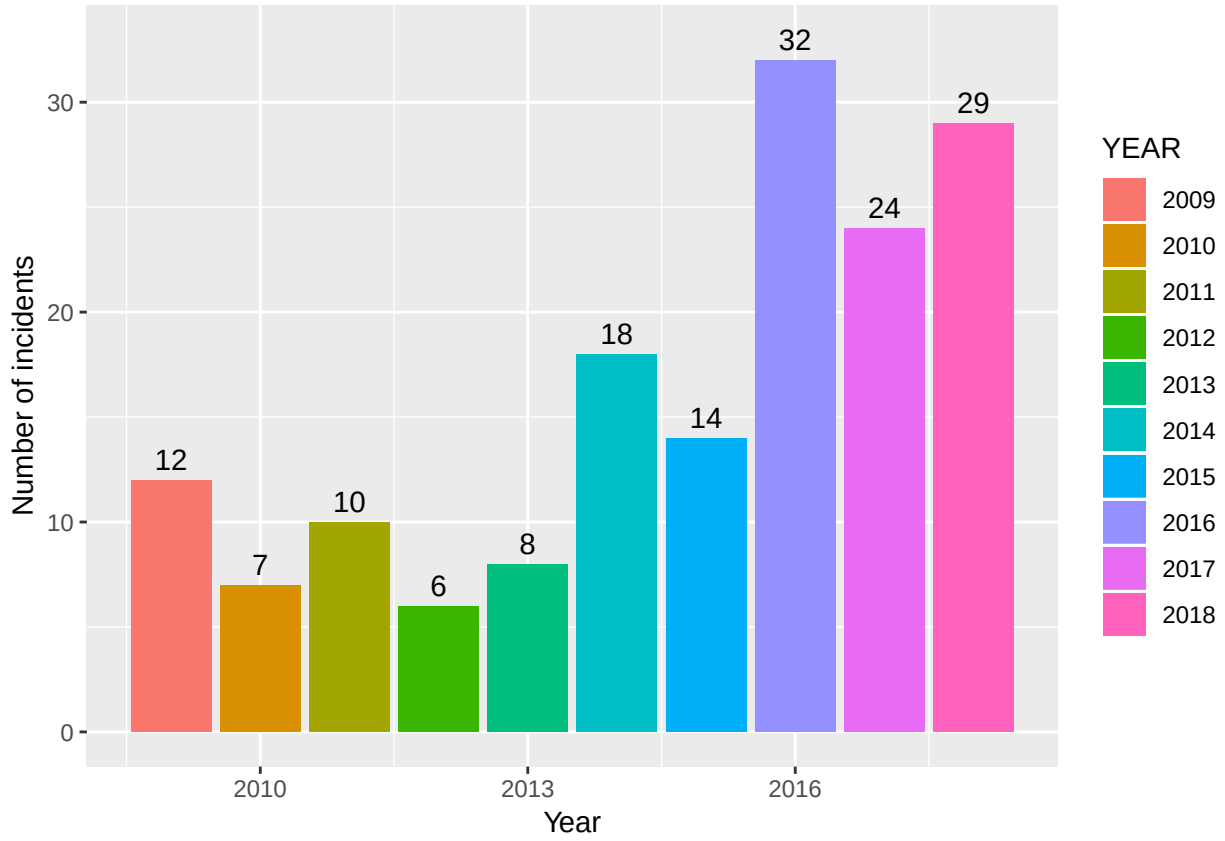
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
main
```

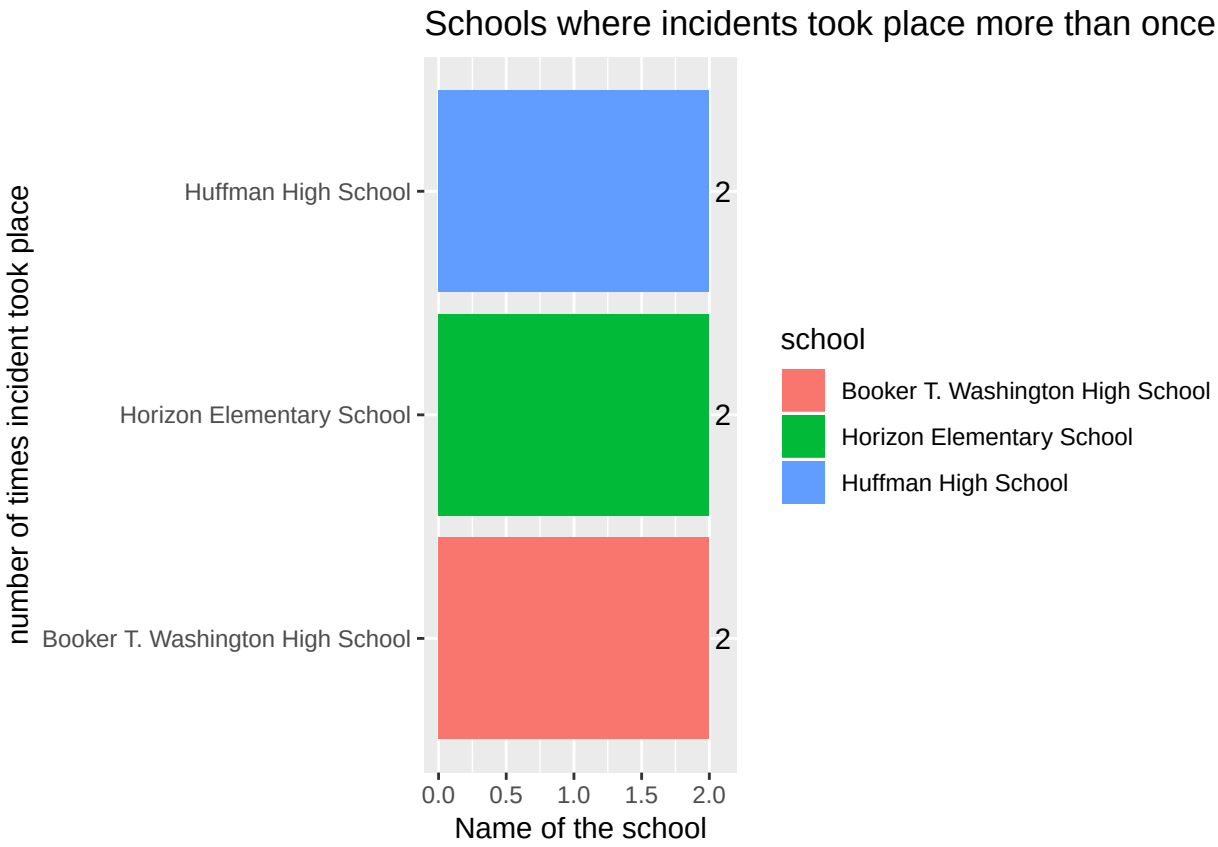
```
## # A tibble: 160 x 17
##    LONCOD LATCOD CDCODE  year school city  state address urbanrural race
##     <dbl>  <dbl>  <dbl> <dbl> <chr>  <chr> <chr> <chr>   <chr>      <chr>
##  1  -75.6   39.7   1000  2009 Willi~ New ~ Dela~ 713 E ~ suburbtown BLACK
##  2  -87.6   41.8   1701  2009 Dunba~ Chic~ Illi~ 3000 S~ urban      BLACK
##  3  -90.2   38.6   1712  2009 Cahok~ Caho~ Illi~ 800 Ra~ suburbtown BLACK
##  4  -78.3   35.8   3713  2009 Zebul~ Zebu~ Nort~ Halifa~ suburbtown BLACK
##  5  -79.0   35.1   3702  2009 Westo~ Faye~ Nort~ 277 Bo~ urban      BLACK
##  6  -77.4   37.5   5103  2009 Chimb~ Rich~ Virg~ 3000 E~ urban      BLACK
##  7  -84.5   38.1   2106  2009 Leest~ Lexi~ Kent~ 2010 L~ urban      BLACK
##  8  -92.9   42.6   1901  2009 Aplin~ Park~ Iowa  610 N ~ rural      WHITE
##  9  -72.5   41.0   3601  2009 Matti~ Matt~ New ~ 15125 ~ suburbtown WHITE
## 10 -118.    33.8    637  2009 Wilso~ Long~ Cali~ 4400 E~ urban      HISP
## # ... with 150 more rows, and 7 more variables: killed <dbl>,
## #   injured <dbl>, victims <dbl>, date <chr>, type <chr>, NCESSCH <dbl>,
## #   time <time>
```

```
#YEAR BAR GRAPH
main %>% group_by(year) %>% summarise(count=n()) %>% ggplot(mapping = aes(x=year,y=count,fill=as.factor
```



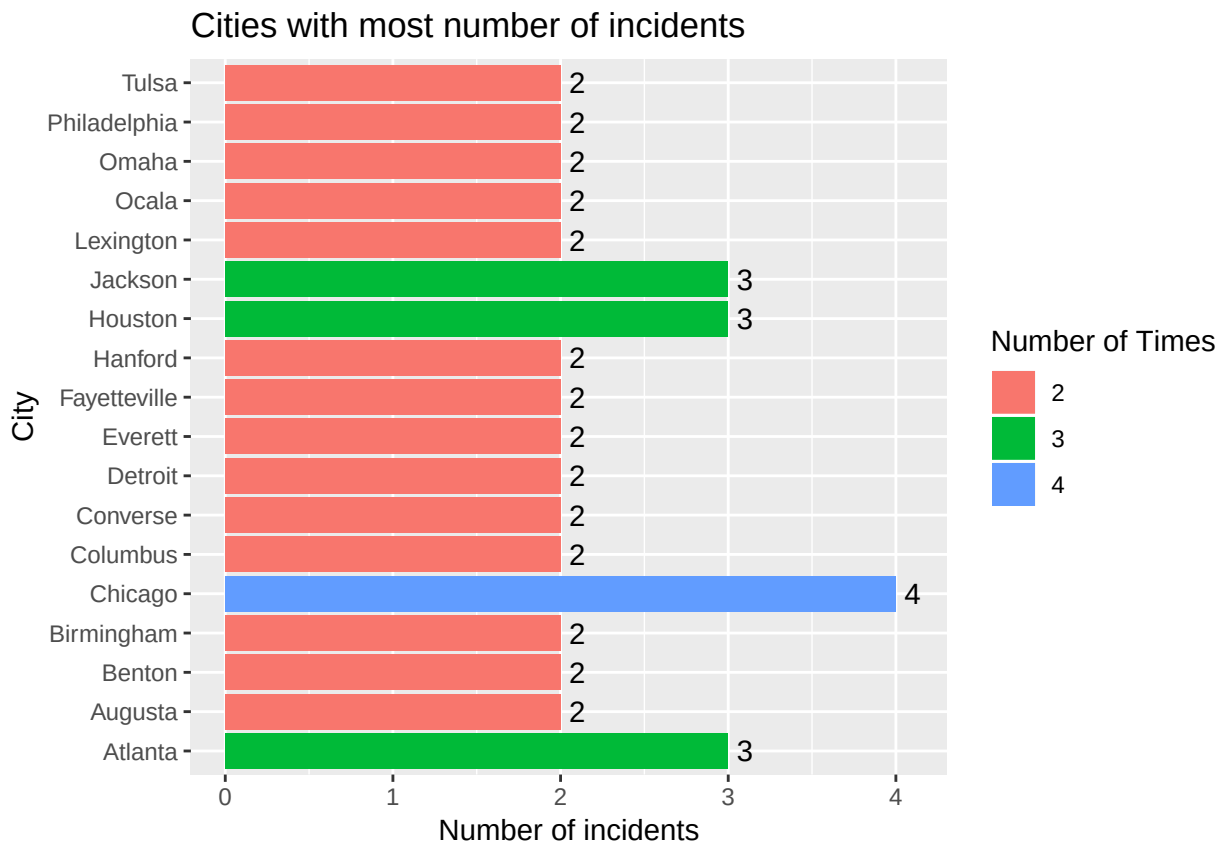

```
main %>% group_by(school) %>% summarise(times=n()) %>% arrange(desc(times))%>% top_n(3) %>% ggplot(mapp
```

```
## Selecting by times
```



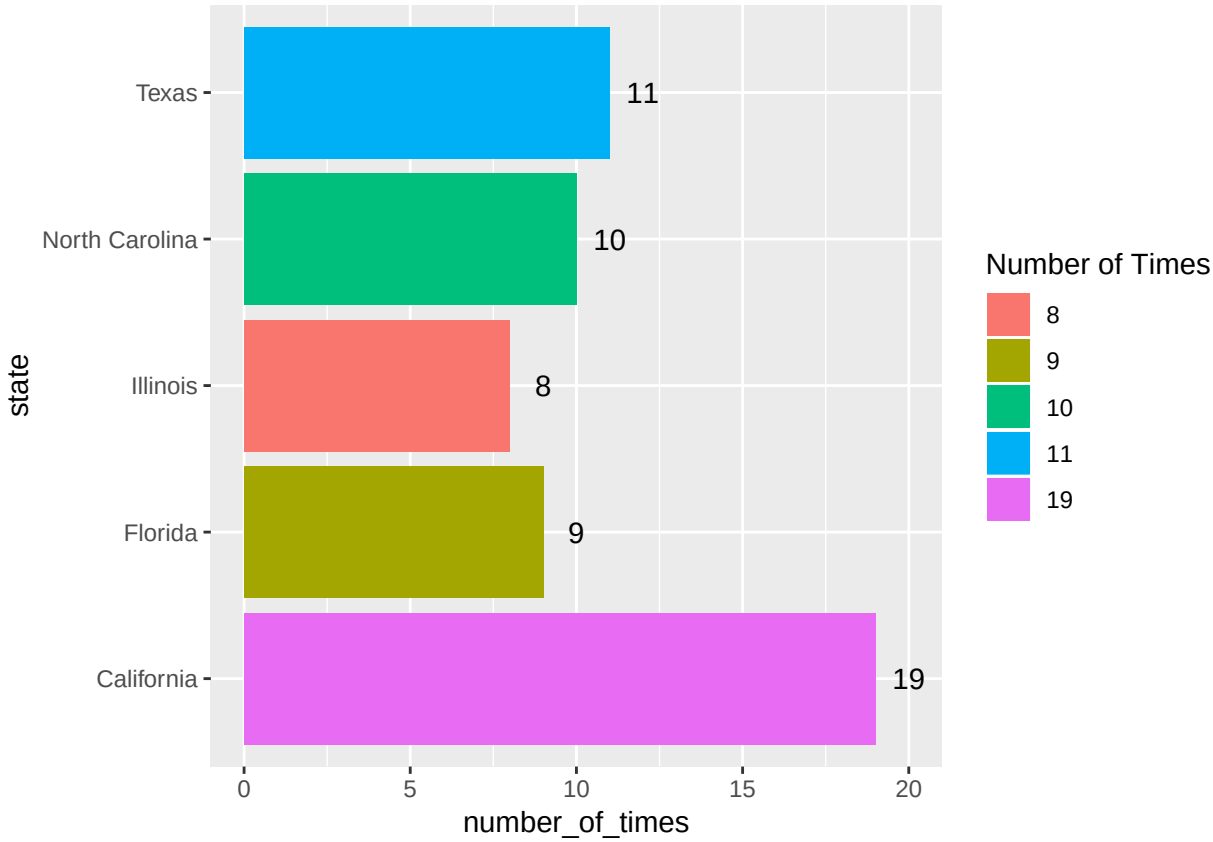

```
#CITIES
main %>% group_by(city) %>% summarise(number_of_times=n()) %>% arrange(desc(number_of_times))%>% top_n(5
```

```
## Selecting by number_of_times
```



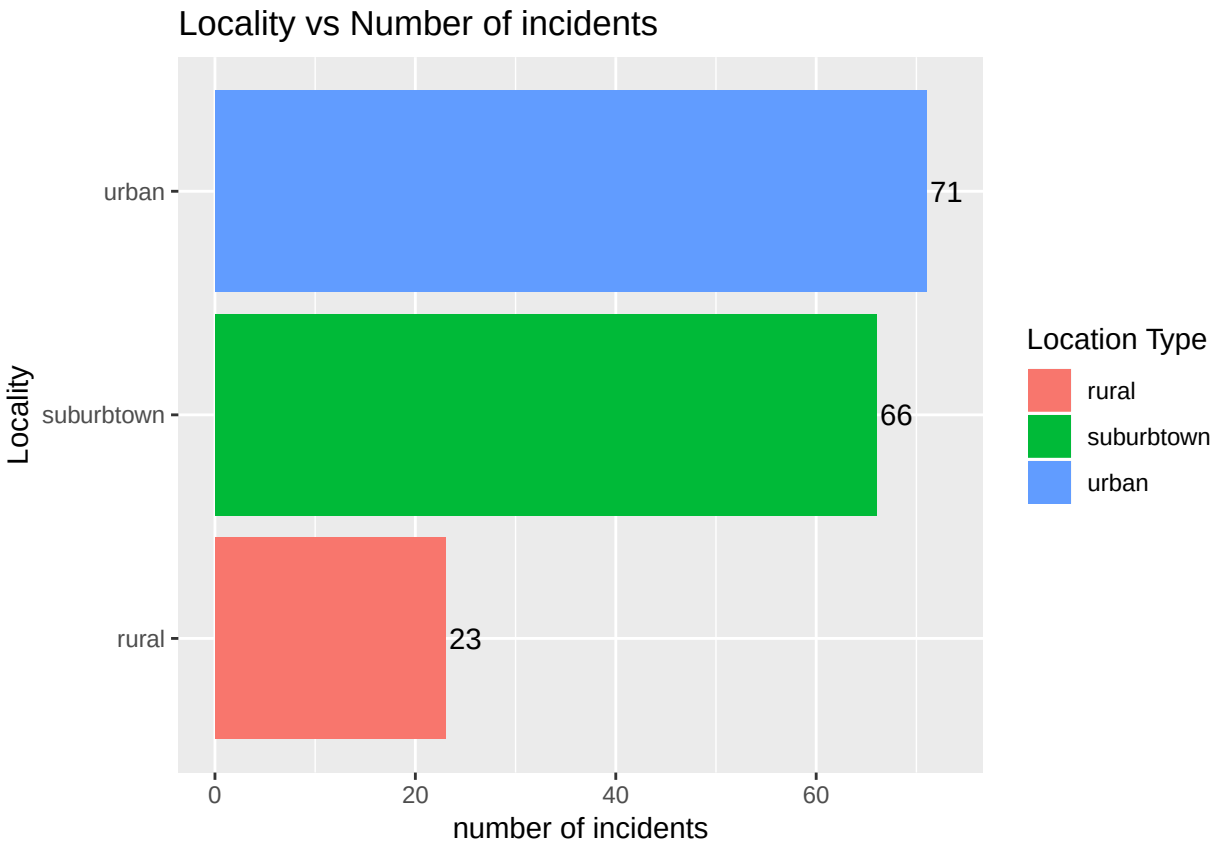

```
#STATES
main %>% group_by(state) %>% summarise(number_of_times=n()) %>% arrange(desc(number_of_times))%>% top_n
```

```
## Selecting by number_of_times
```



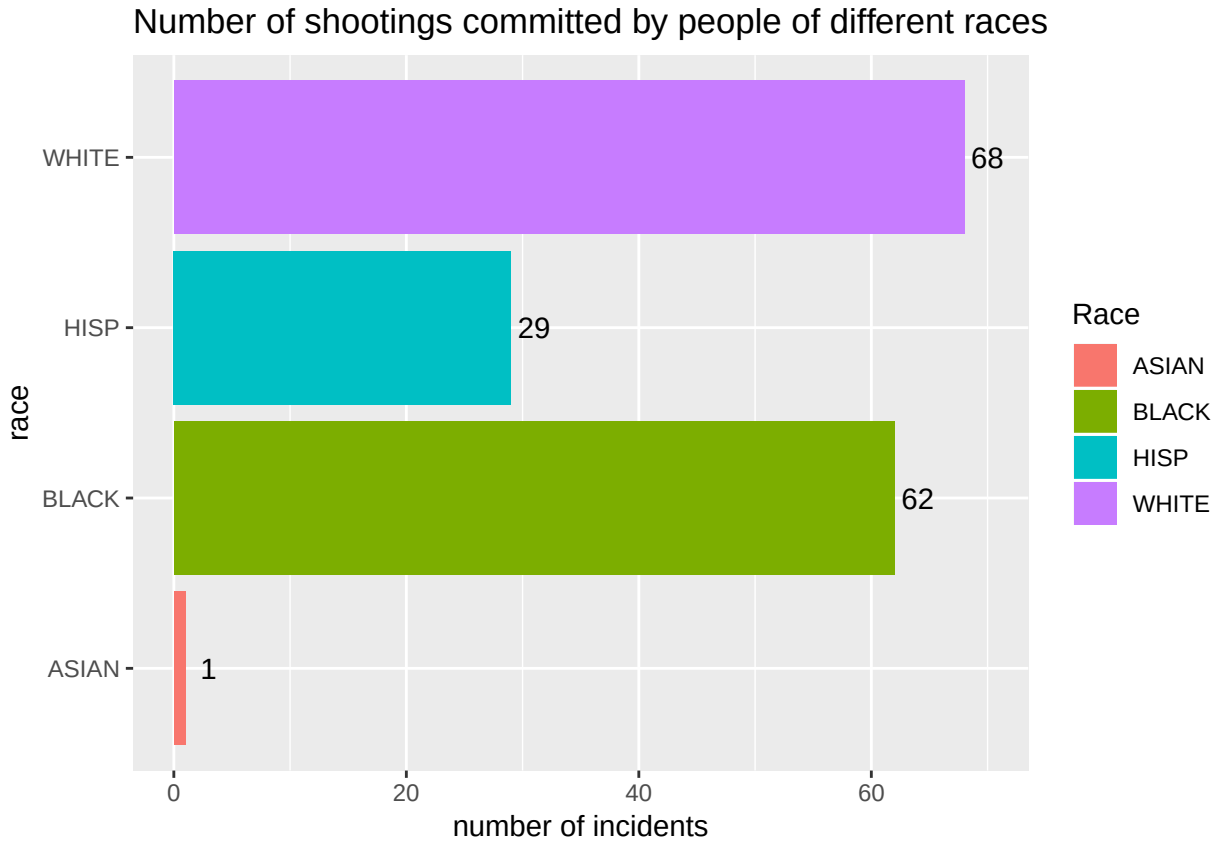

```
#urbanrural
main %>% group_by(urbanrural) %>% summarise(number=n()) %>% arrange(desc(number))%>% top_n(5) %>% ggplo
```

```
## Selecting by number
```



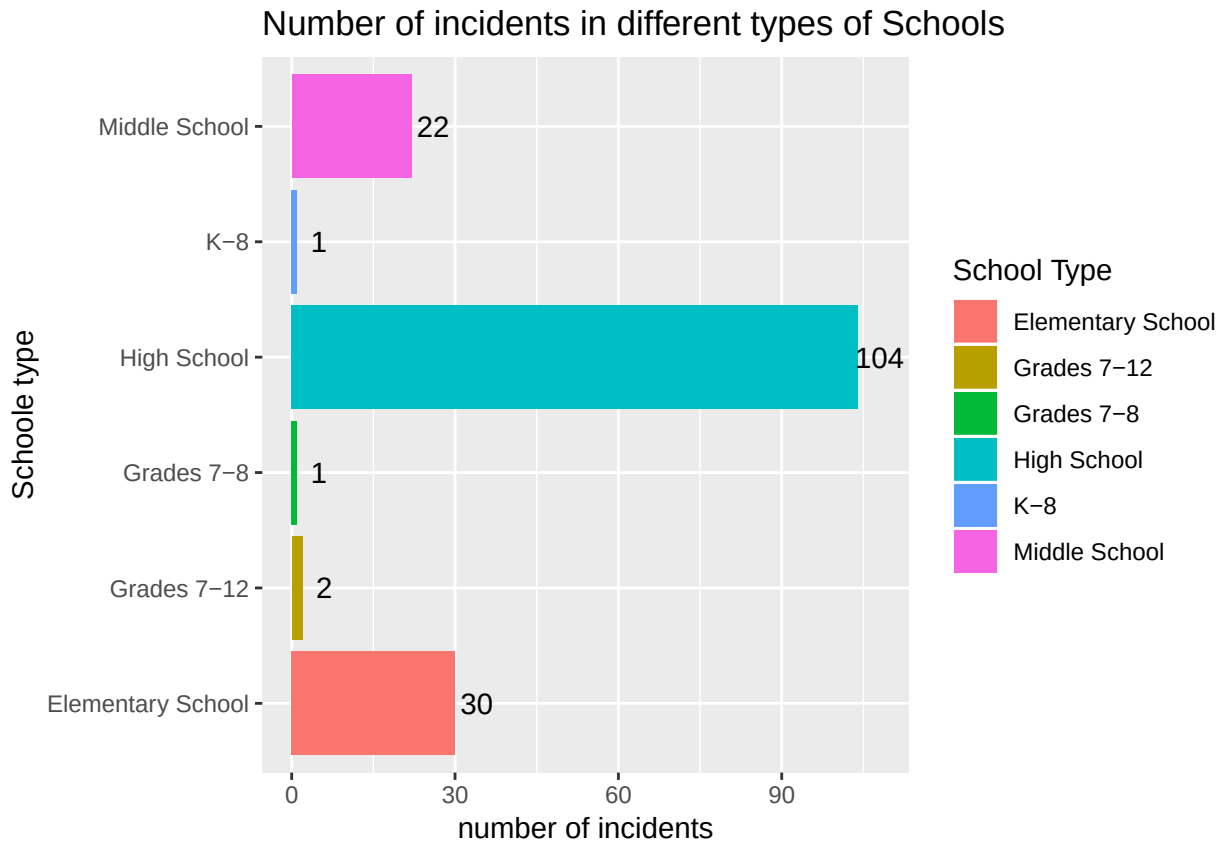

```
#Race
main %>% group_by(race) %>% summarise(counting=n()) %>% arrange(desc(counting))%>% top_n(5) %>% ggplot(
```

```
## Selecting by counting
```




```
#SCHOOL TYPE
main %>% group_by(type) %>% summarise(total=n()) %>% arrange(desc(total))%>% top_n(5) %>% ggplot(mapping
```

```
## Selecting by total
```

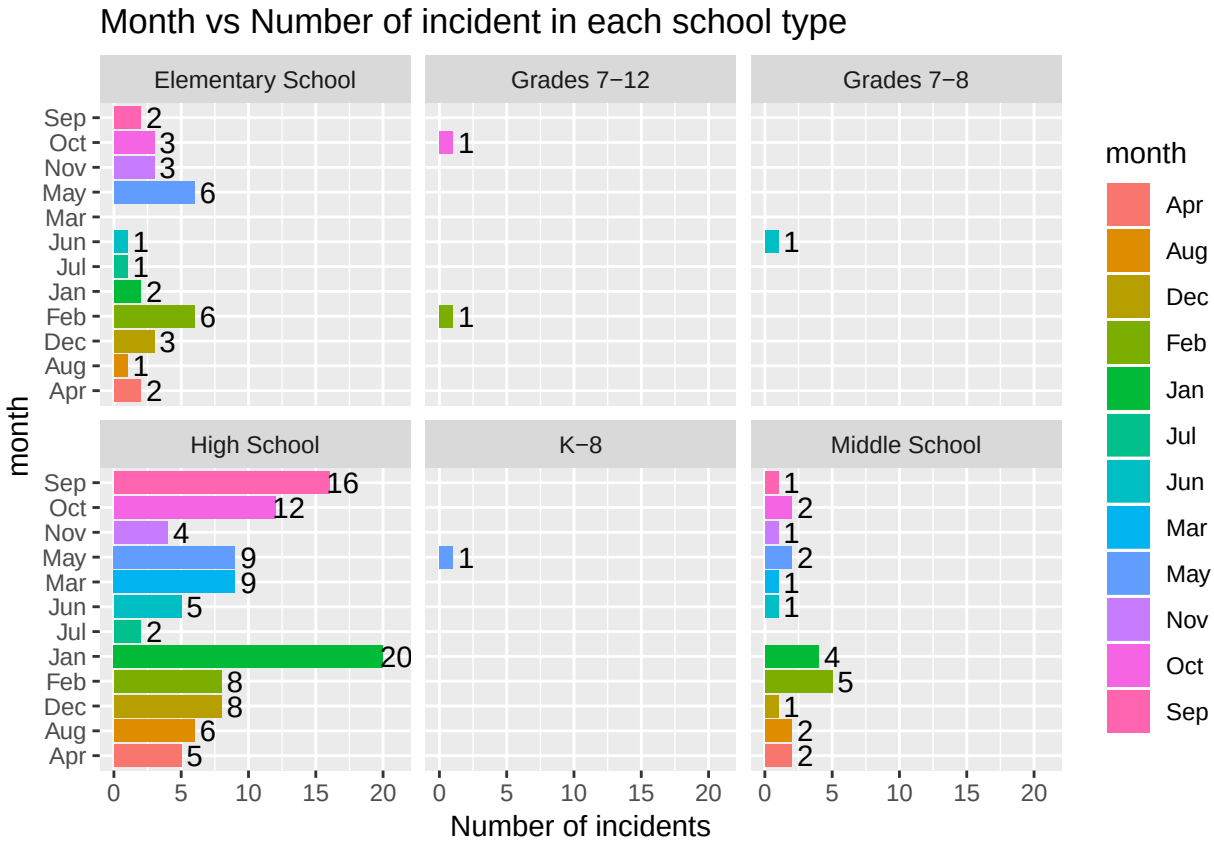



```
main<-main %>% mutate(month=str_sub(main[['date']],-6,-4))
main
```

```
## # A tibble: 160 x 18
##    LONCOD LATCOD CDCODE  year school city  state address urbanrural race
##     <dbl>  <dbl>  <dbl> <dbl> <chr>  <chr> <chr> <chr>   <chr>      <chr>
##  1  -75.6   39.7   1000  2009 Willi~ New ~ Dela~ 713 E ~ suburbtown BLACK
##  2  -87.6   41.8   1701  2009 Dunba~ Chic~ Illi~ 3000 S~ urban      BLACK
##  3  -90.2   38.6   1712  2009 Cahok~ Caho~ Illi~ 800 Ra~ suburbtown BLACK
##  4  -78.3   35.8   3713  2009 Zebul~ Zebu~ Nort~ Halifa~ suburbtown BLACK
##  5  -79.0   35.1   3702  2009 Westo~ Faye~ Nort~ 277 Bo~ urban      BLACK
##  6  -77.4   37.5   5103  2009 Chimb~ Rich~ Virg~ 3000 E~ urban      BLACK
##  7  -84.5   38.1   2106  2009 Leest~ Lexi~ Kent~ 2010 L~ urban      BLACK
##  8  -92.9   42.6   1901  2009 Aplin~ Park~ Iowa  610 N ~ rural      WHITE
##  9  -72.5   41.0   3601  2009 Matti~ Matt~ New ~ 15125 ~ suburbtown WHITE
## 10 -118.    33.8    637  2009 Wilso~ Long~ Cali~ 4400 E~ urban      HISP
## # ... with 150 more rows, and 8 more variables: killed <dbl>,
## #   injured <dbl>, victims <dbl>, date <chr>, type <chr>, NCESSCH <dbl>,
## #   time <time>, month <chr>
```

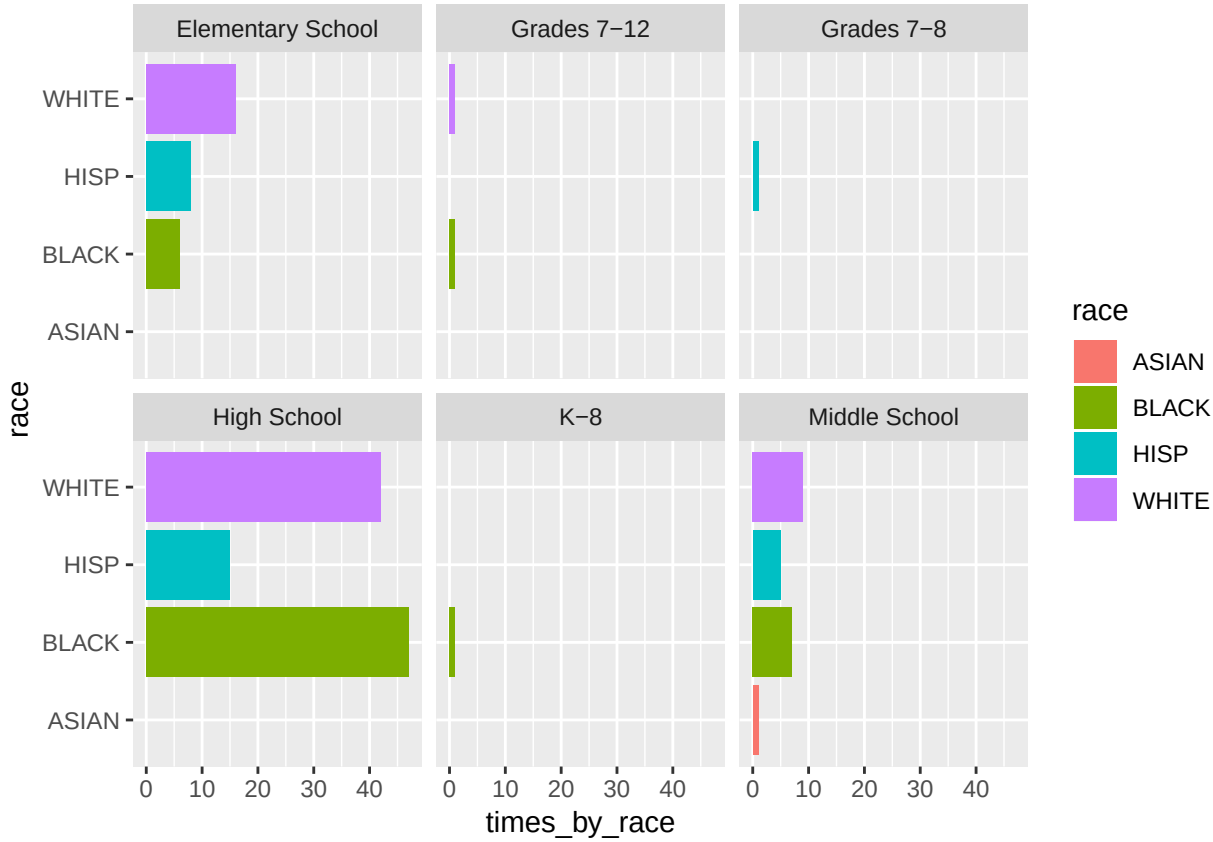
```
#MONTHS EACH SCHOOL TYPE
main %>% group_by(month,type) %>% summarise(times_in_month=n()) %>% arrange(desc(times_in_month)) %>% t
```

```
## Selecting by times_in_month
```



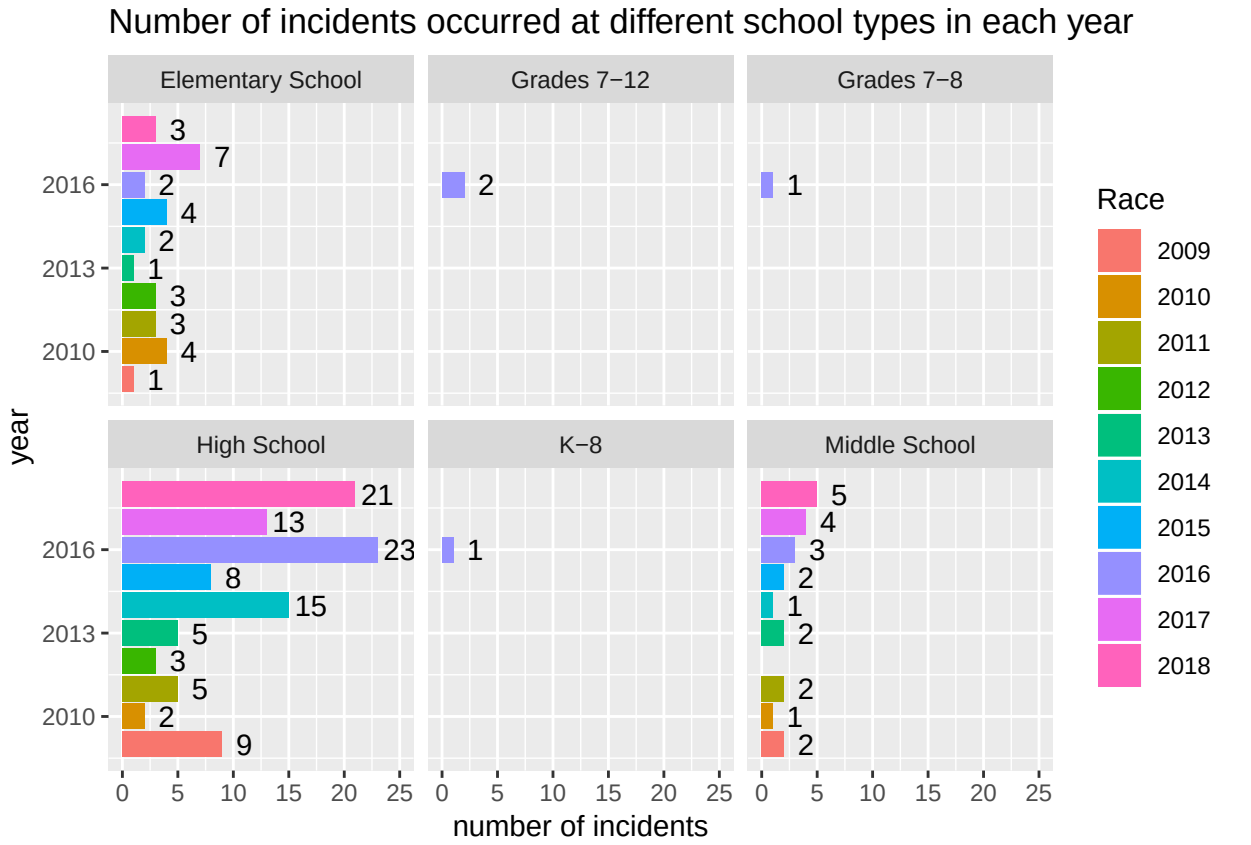

```
#Race in each school type
main %>% group_by(race,type) %>% summarise(times_by_race=n()) %>% arrange(desc(times_by_race)) %>% top_
```

```
## Selecting by times_by_race
```



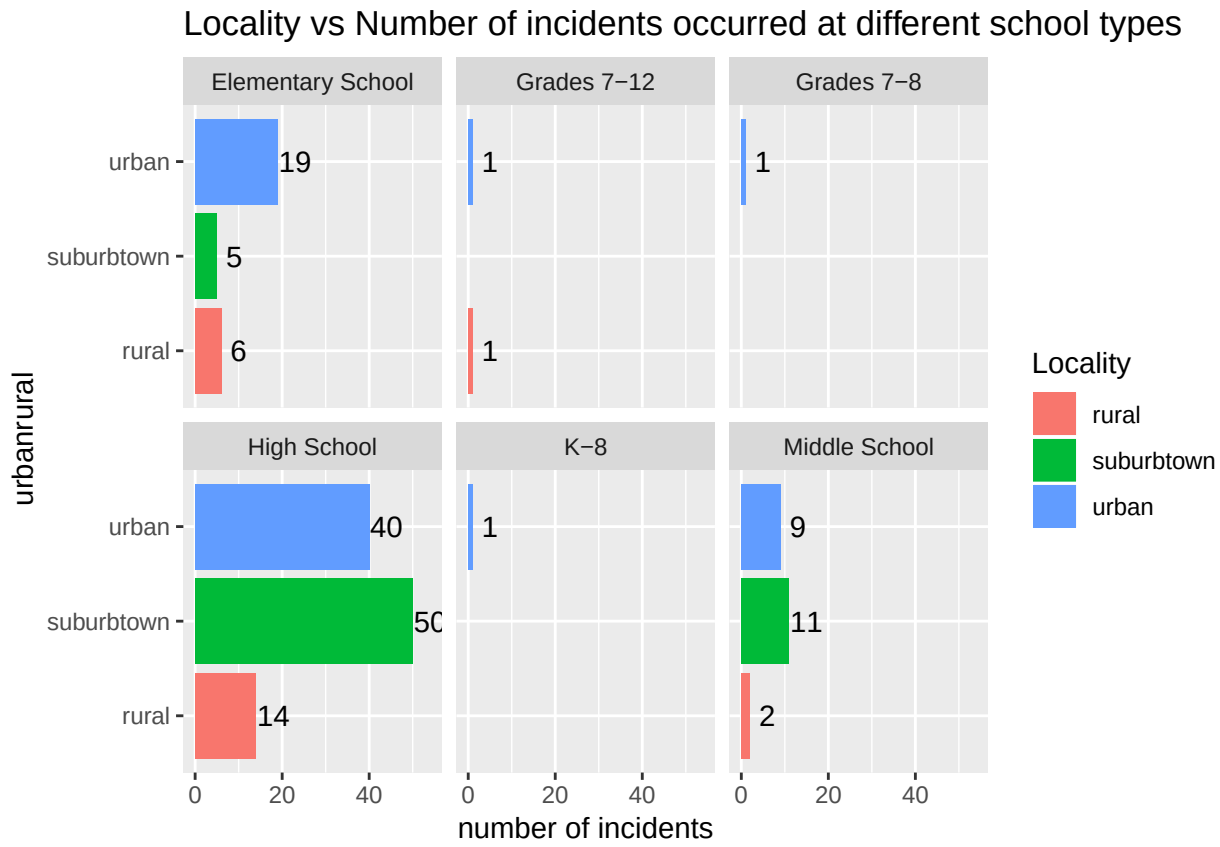

```
#years in school type
main %>% group_by(year,type) %>% summarise(times_in_year=n()) %>% arrange(desc(times_in_year)) %>% top_
```

```
## Selecting by times_in_year
```



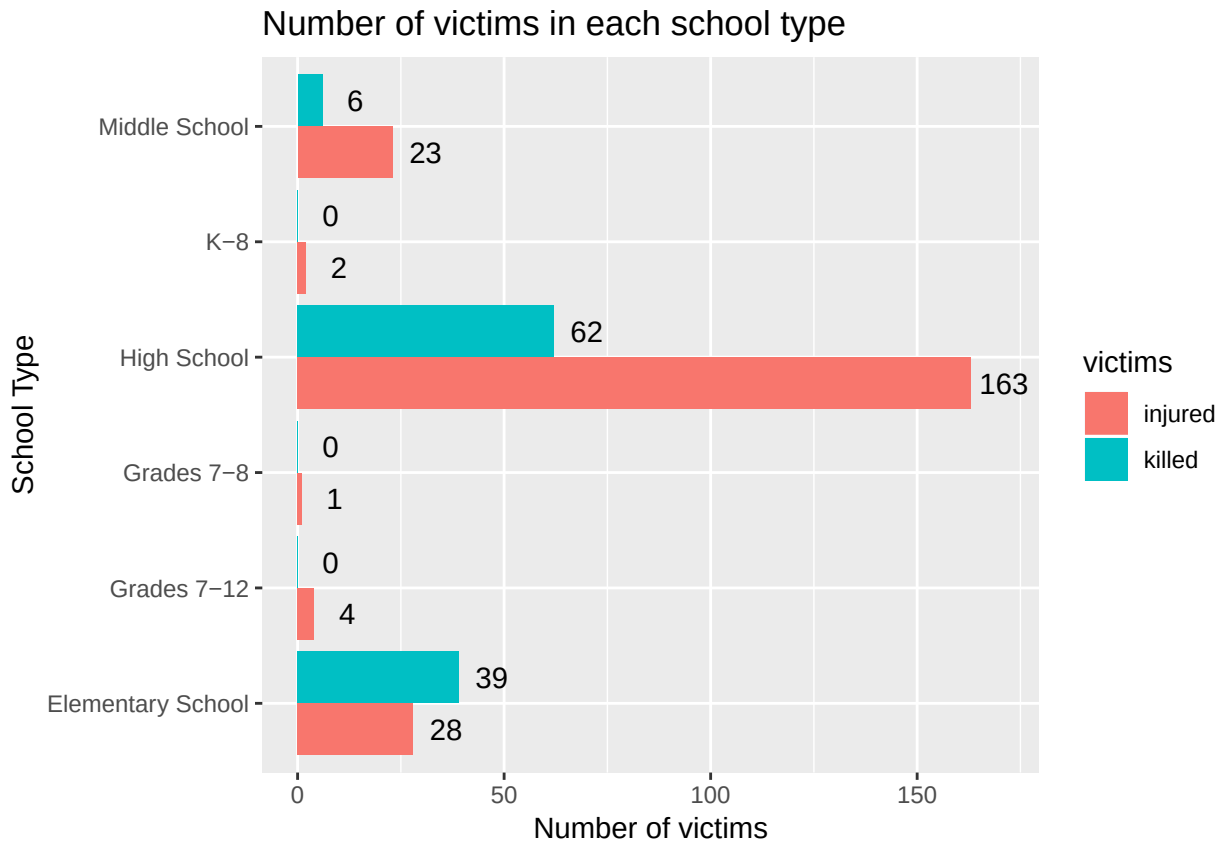

```
 # urbanrural vs school type
main %>% group_by(urbanrural,type) %>% summarise(number=n()) %>% arrange(desc(number)) %>% top_n(5) %>%
```

```
## Selecting by number
```



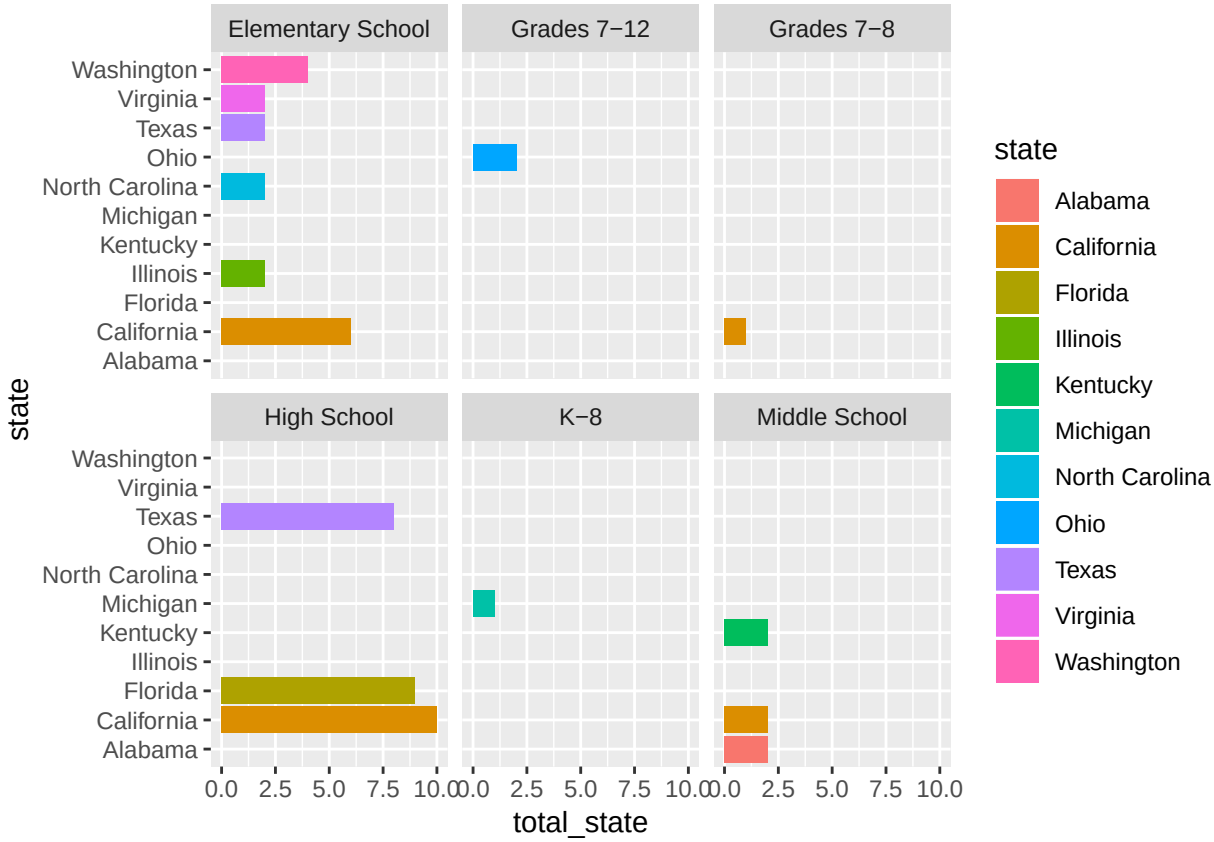

```
#VICTIMS IN EACH SCHOOL TYPE
main %>% gather(key='victims',value='victims_count',killed,injured) %>% group_by(type,victims) %>% summa
```

```
## Selecting by total_victims
```



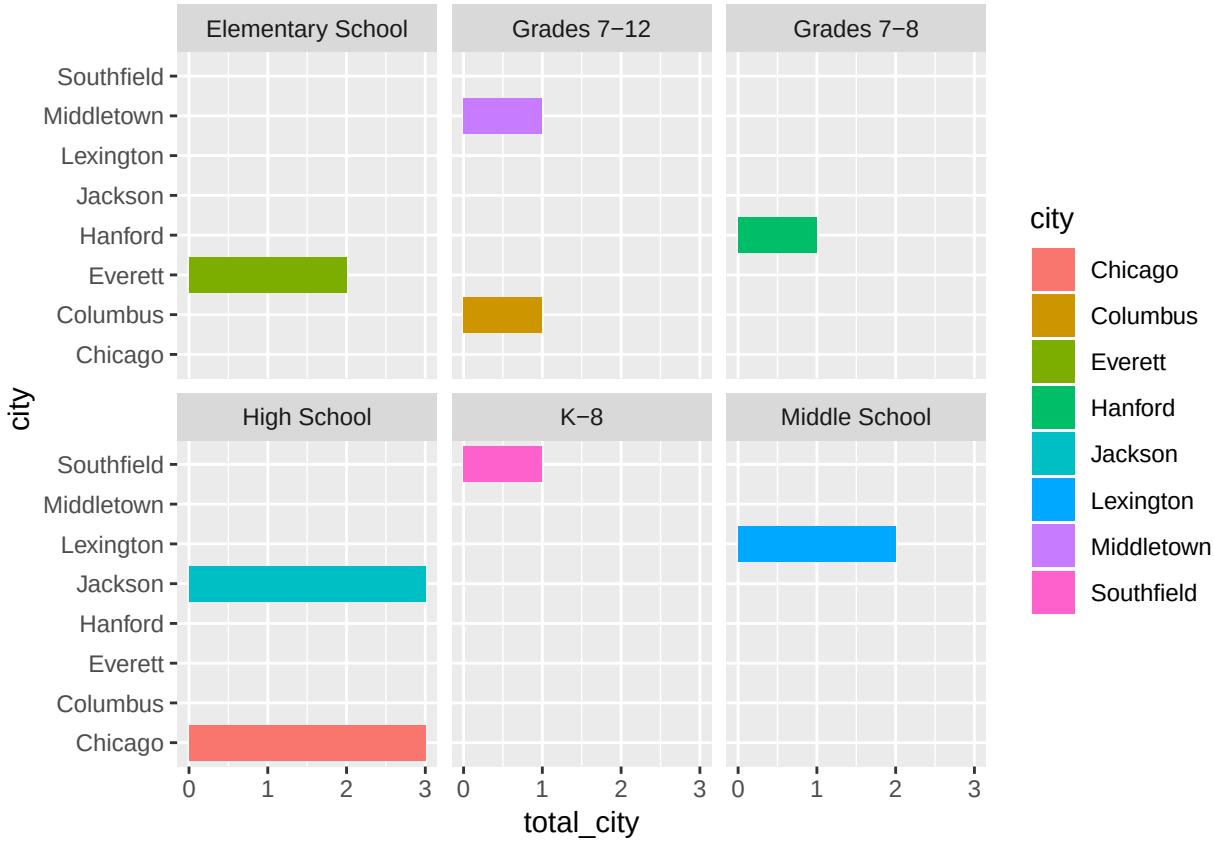

```
#state vs type
main %>% group_by(state,type) %>% summarise(total_state=n()) %>% arrange(desc(total_state)) %>% group_by
```

```
## Selecting by total_state
```




```
#CITY VS TYPE
main %>% group_by(city,type) %>% summarise(total_city=n()) %>% arrange(desc(total_city)) %>% group_by(ty
```

```
## Selecting by total_city
```




```
one<-main %>% gather(key='victims',value='victims_count',killed,injured)
```