

Project

Load Libraries

```
library(sqldf)  
  
## Loading required package: gsubfn  
  
## Loading required package: proto  
  
## Loading required package: RSQLite  
  
library(ggplot2)  
library(reshape2)
```

Load Project Data

```
# Load data  
my_mv <- read.csv("~/01 Personal/MS/IST 687/Project/Motor_Vehicle_Crashes_-_Vehicle_Information_  
_Three_Year_Window.csv", stringsAsFactors=FALSE)
```

Prepare Data for YOY Breakdown Analysis on various categories

```

my_veh_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`Vehicle.Body.Type` ,count(1) cnt from my_mv group by year,`Vehicle.Body.Type` )x inner join (select `Vehicle.Body.Type` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `Vehicle.Body.Type` )y on x.`Vehicle.Body.Type` =y.`Vehicle.Body.Type` order by yr_cnt desc,tot_cnt desc')

my_reg_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`Registration.Class` ,count(1) cnt from my_mv group by year,`Registration.Class` )x inner join (select `Registration.Class` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `Registration.Class` )y on x.`Registration.Class` =y.`Registration.Class` order by yr_cnt desc,tot_cnt desc')

my_act_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`Action.Prior.to.Accident` ,count(1) cnt from my_mv group by year,`Action.Prior.to.Accident` )x inner join (select `Action.Prior.to.Accident` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `Action.Prior.to.Accident` )y on x.`Action.Prior.to.Accident` =y.`Action.Prior.to.Accident` order by yr_cnt desc,tot_cnt desc')

my_dir_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`Direction.of.Travel` ,count(1) cnt from my_mv group by year,`Direction.of.Travel` )x inner join (select `Direction.of.Travel` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `Direction.of.Travel` )y on x.`Direction.of.Travel` =y.`Direction.of.Travel` order by yr_cnt desc,tot_cnt desc')

my_fuel_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`Fuel.Type` ,count(1) cnt from my_mv group by year,`Fuel.Type` )x inner join (select `Fuel.Type` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `Fuel.Type` )y on x.`Fuel.Type` =y.`Fuel.Type` order by yr_cnt desc,tot_cnt desc')

my_st_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`State.of.Registration` ,count(1) cnt from my_mv group by year,`State.of.Registration` )x inner join (select `State.of.Registration` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `State.of.Registration` )y on x.`State.of.Registration` =y.`State.of.Registration` order by yr_cnt desc,tot_cnt desc')

my_make_dist<-sqldf(' select x.* ,y.tot_cnt,y.yr_cnt from (select year,`Vehicle.Make` ,count(1) cnt from my_mv group by year,`Vehicle.Make` )x inner join (select `Vehicle.Make` ,count(1) tot_cnt,count(distinct year) yr_cnt from my_mv group by `Vehicle.Make` )y on x.`Vehicle.Make` =y.`Vehicle.Make` order by yr_cnt desc,tot_cnt desc')

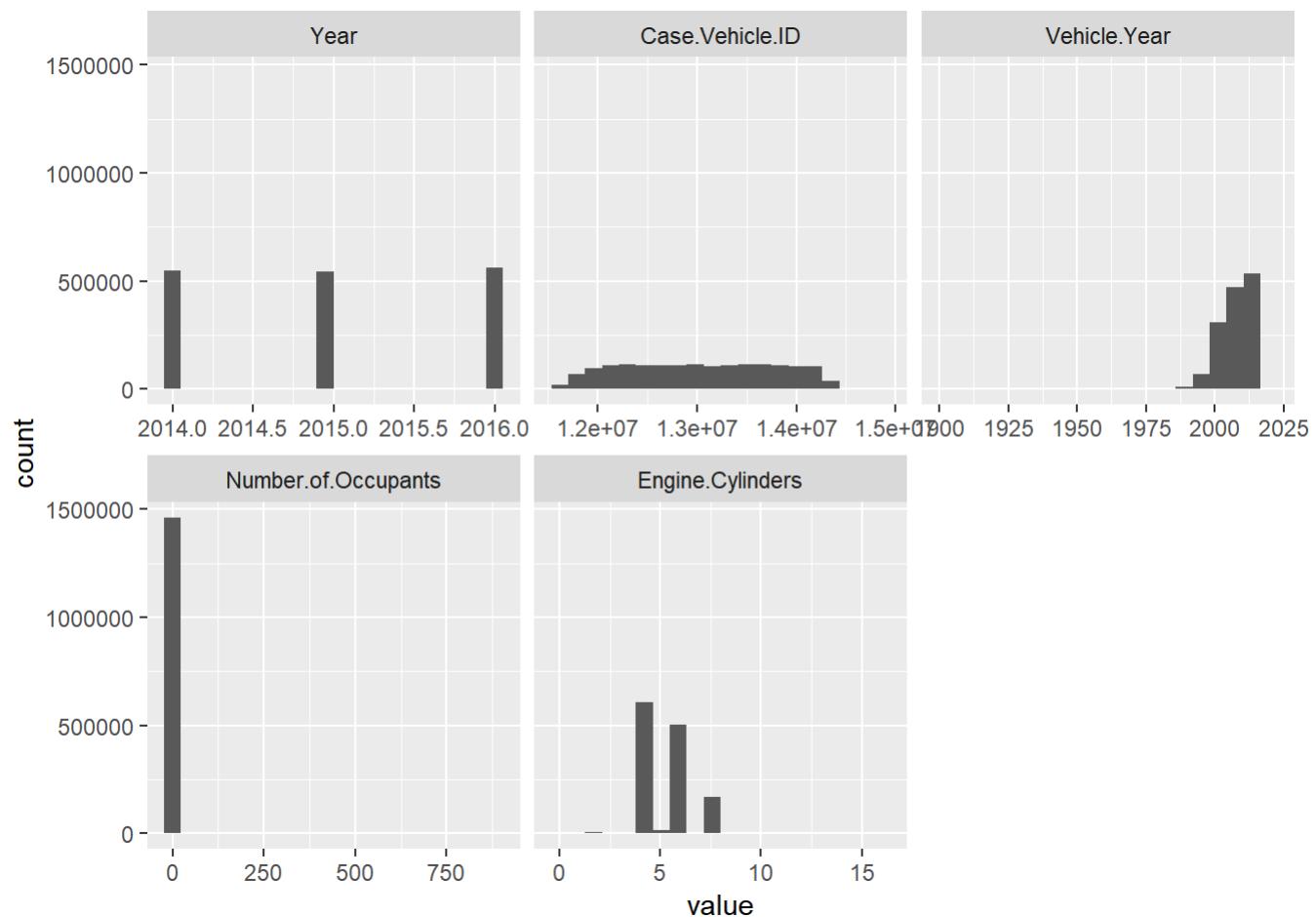
```

Histogram on Numeric metrics available in the project data

```
ggplot(data = melt(my_mv), mapping = aes(x = value)) + geom_histogram(bins = 20) + facet_wrap(~variable, scales = 'free_x')
```

```
## Using Vehicle.Body.Type, Registration.Class, Action.Prior.to.Accident, Type...Axles.of.Truck.or.Bus, Direction.of.Travel, Fuel.Type, State.of.Registration, Vehicle.Make, Contributing.Factor.1, Contributing.Factor.1.Description, Contributing.Factor.2, Contributing.Factor.2.Description, Event.Type, Partial.VIN as id variables
```

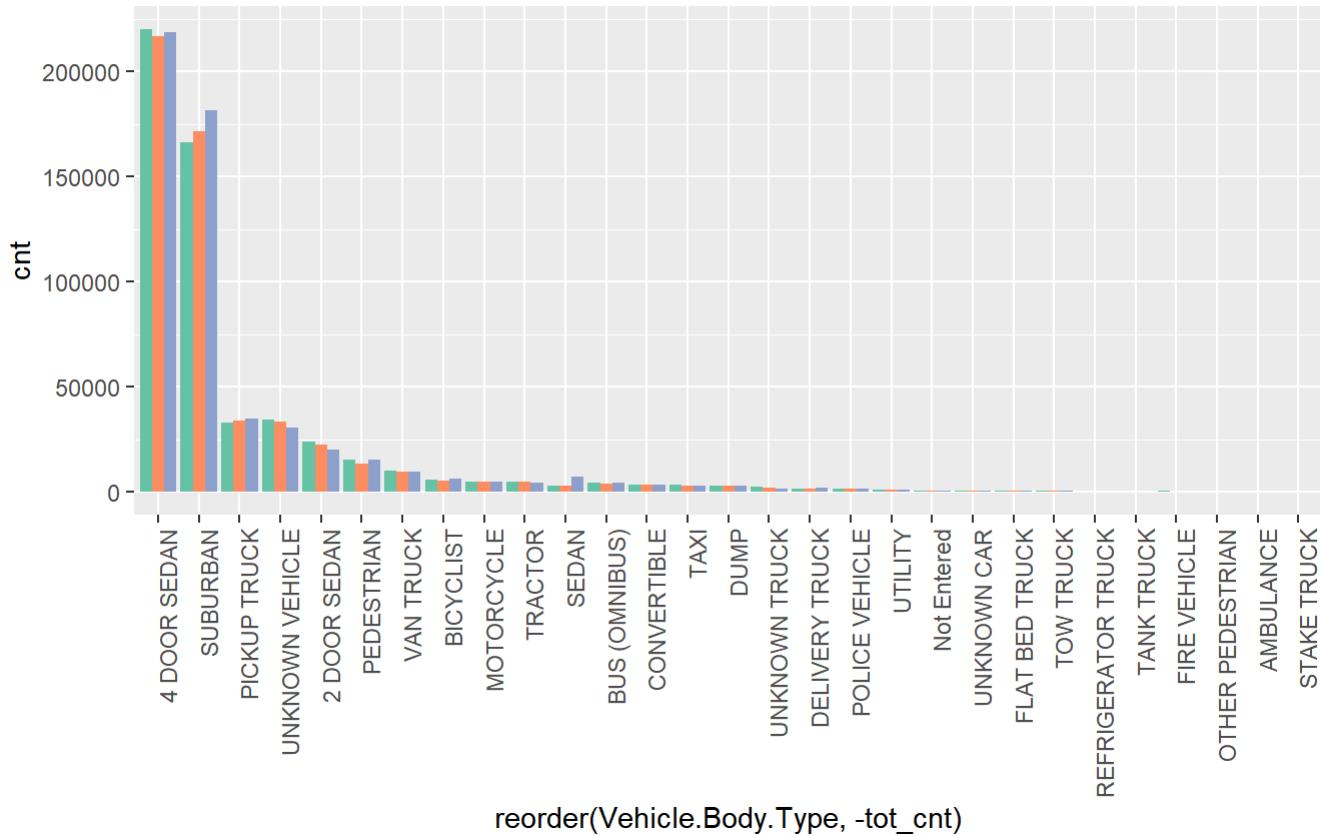
```
## Warning: Removed 796067 rows containing non-finite values (stat_bin).
```



Plot Vehicle.Body.Type YOY breakdown

```
ggplot(my_veh_dist[which(my_veh_dist$tot_cnt>500),],aes(x=reorder(Vehicle.Body.Type,-tot_cnt), y =cnt,fill=as.character(year))) +geom_bar(stat="identity", position = position_dodge(preserve = "total")) + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+ scale_fill_brewer(palette = "Set2")
```

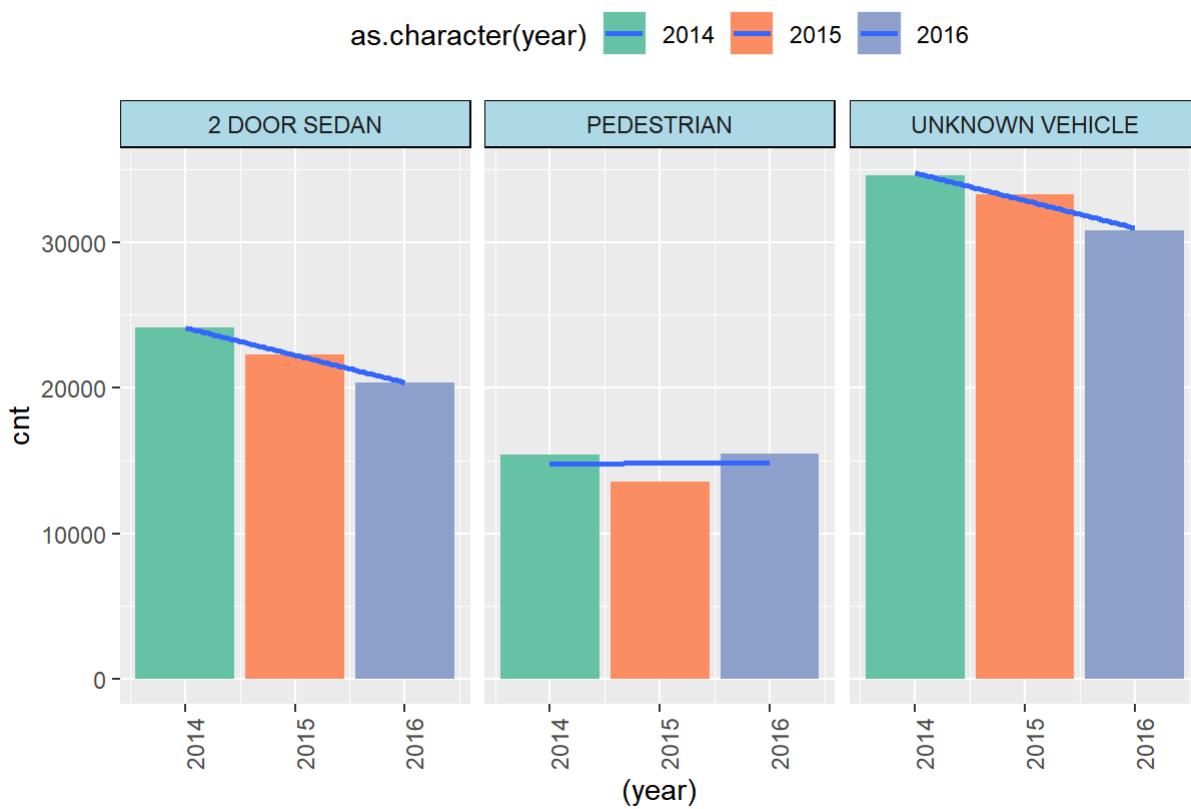
as.character(year) 2014 2015 2016

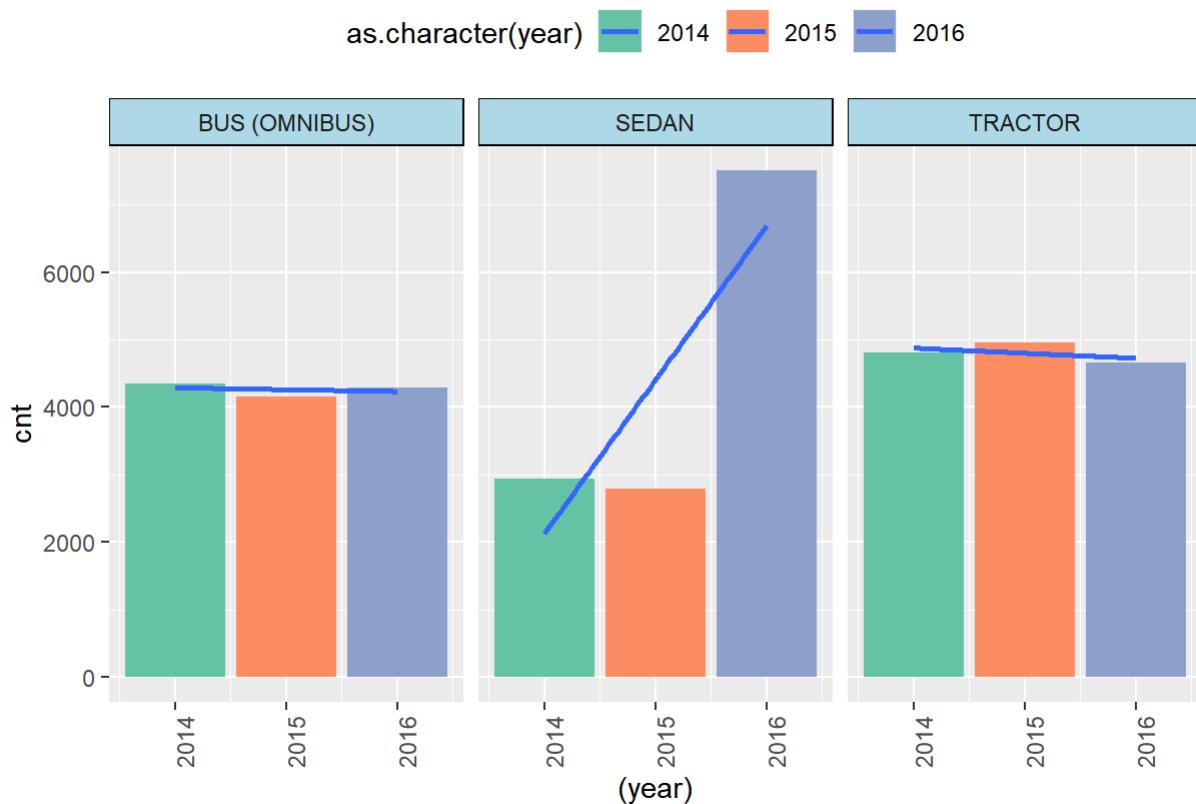
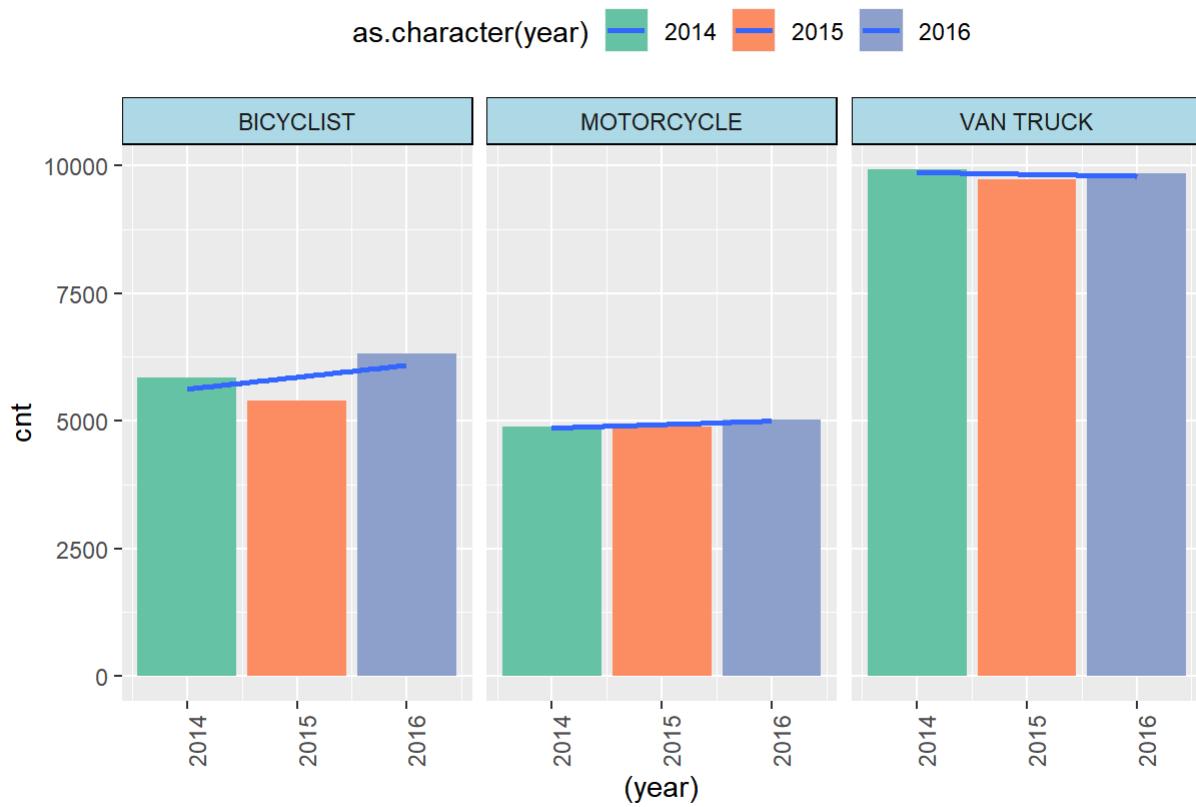


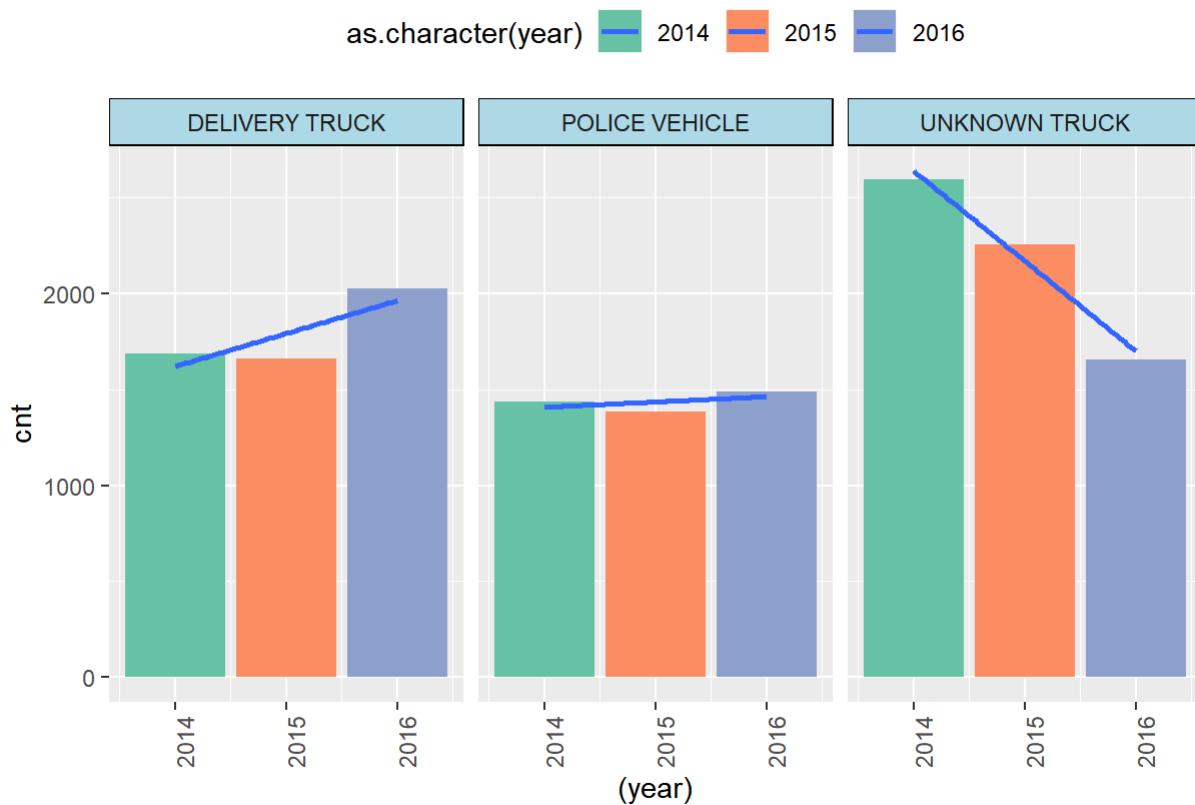
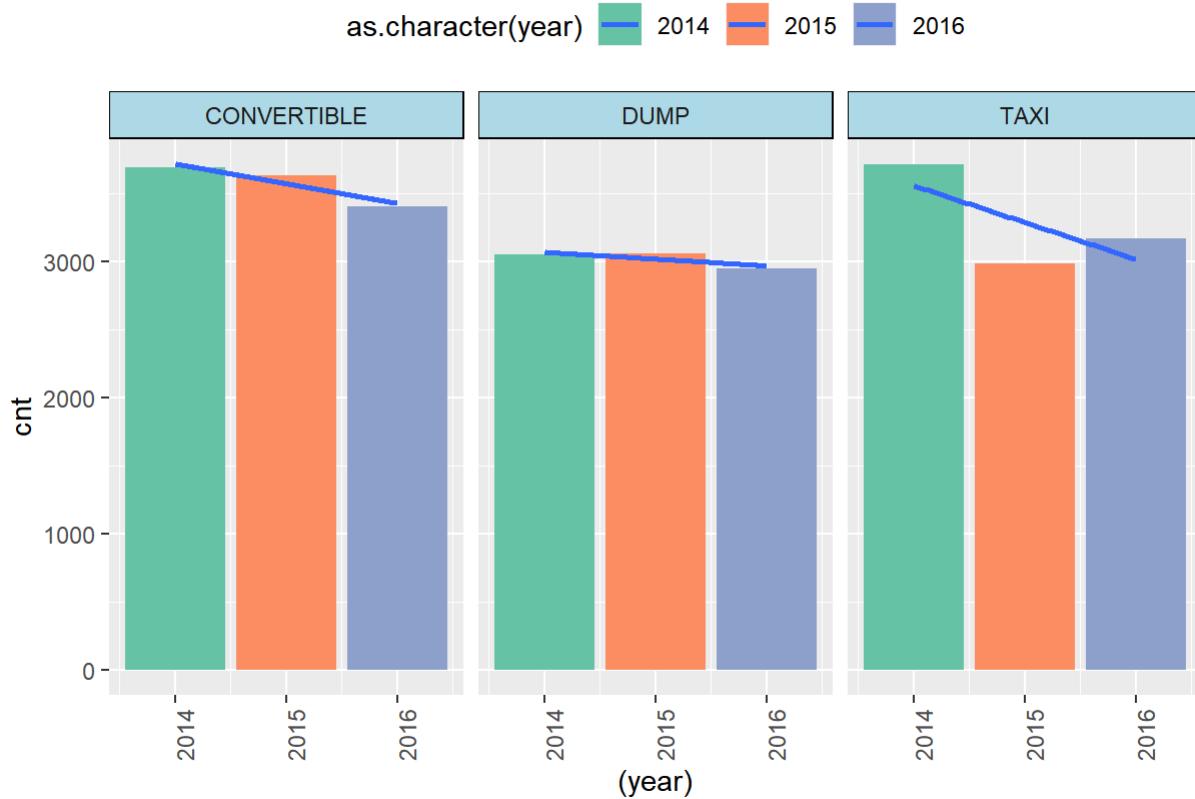
```
#ggplot(my_veh_dist[1:15,],aes(x=(Year), y=cnt,fill=as.character(Year))) +geom_bar(stat="identity") + theme(Legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+ scale_fill_brewer(palette = "Set2") + facet_grid(~Vehicle.Body.Type) + geom_smooth(aes(group = 1))

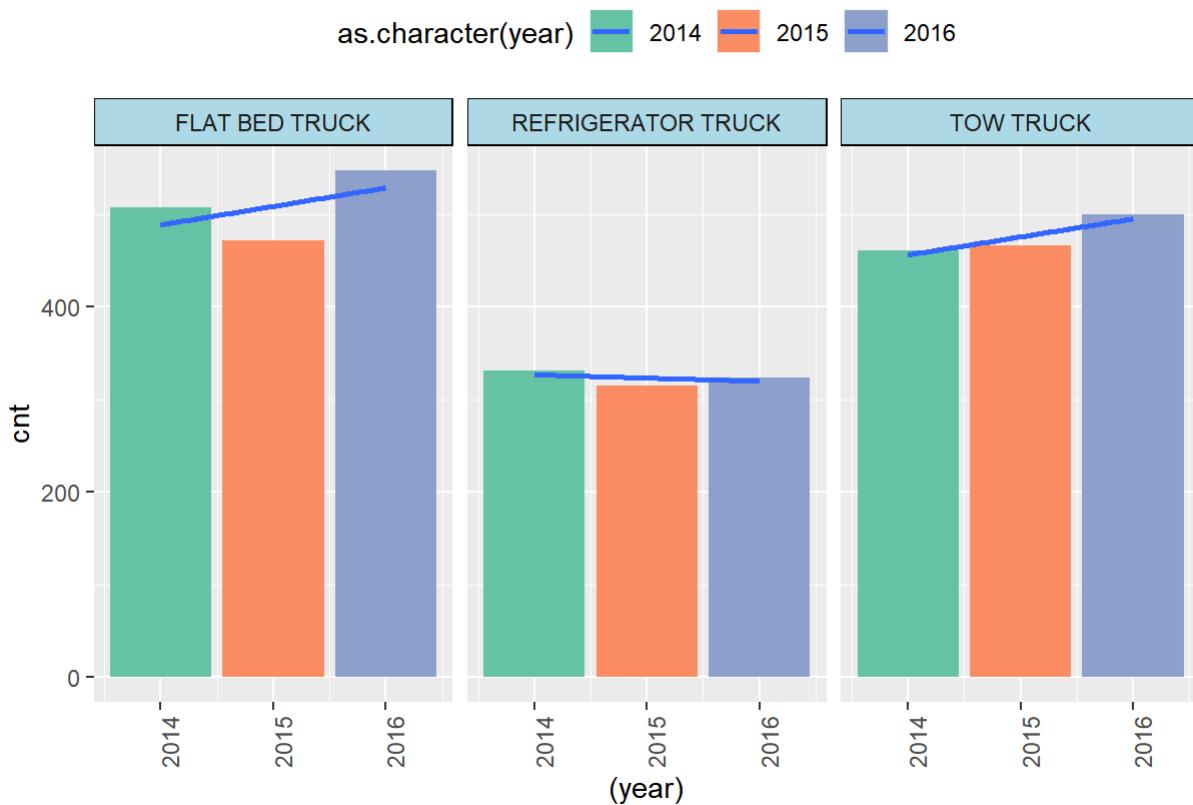
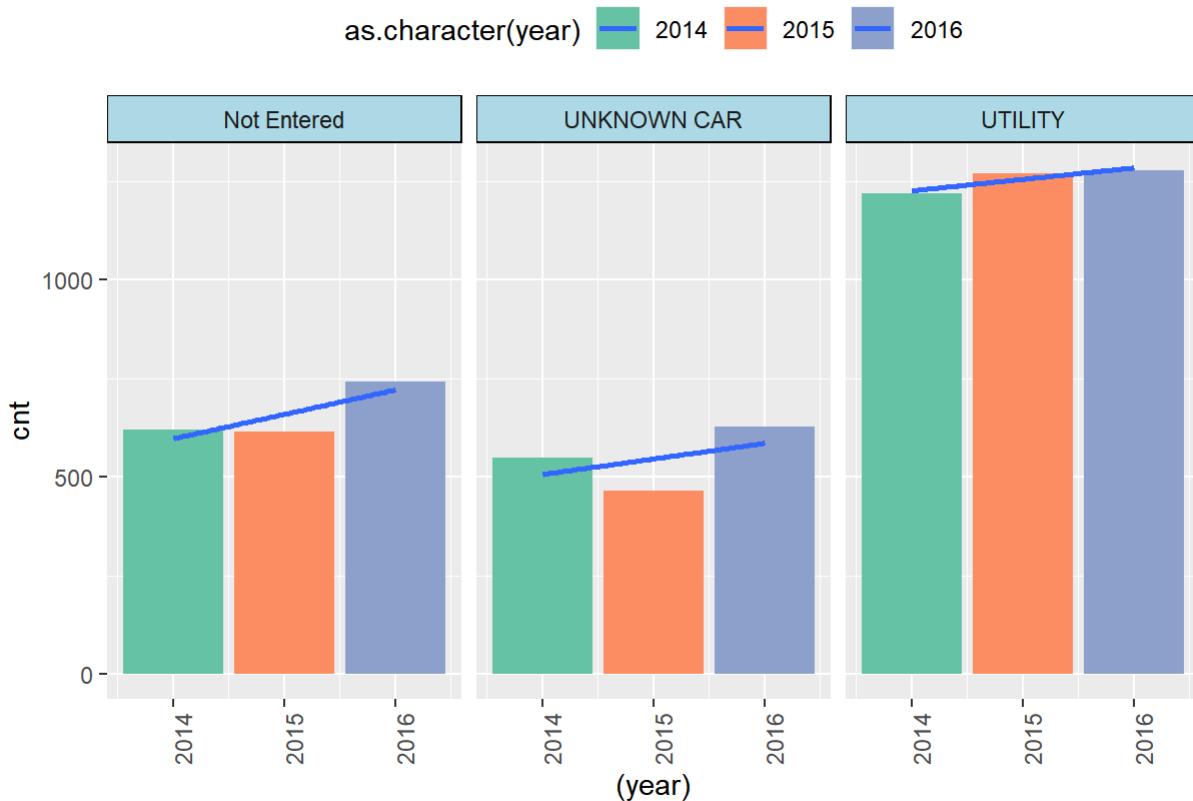
seq <-as.integer(nrow(my_veh_dist[which(my_veh_dist$tot_cnt>300 & my_veh_dist$yr_cnt==3),])/9)+1

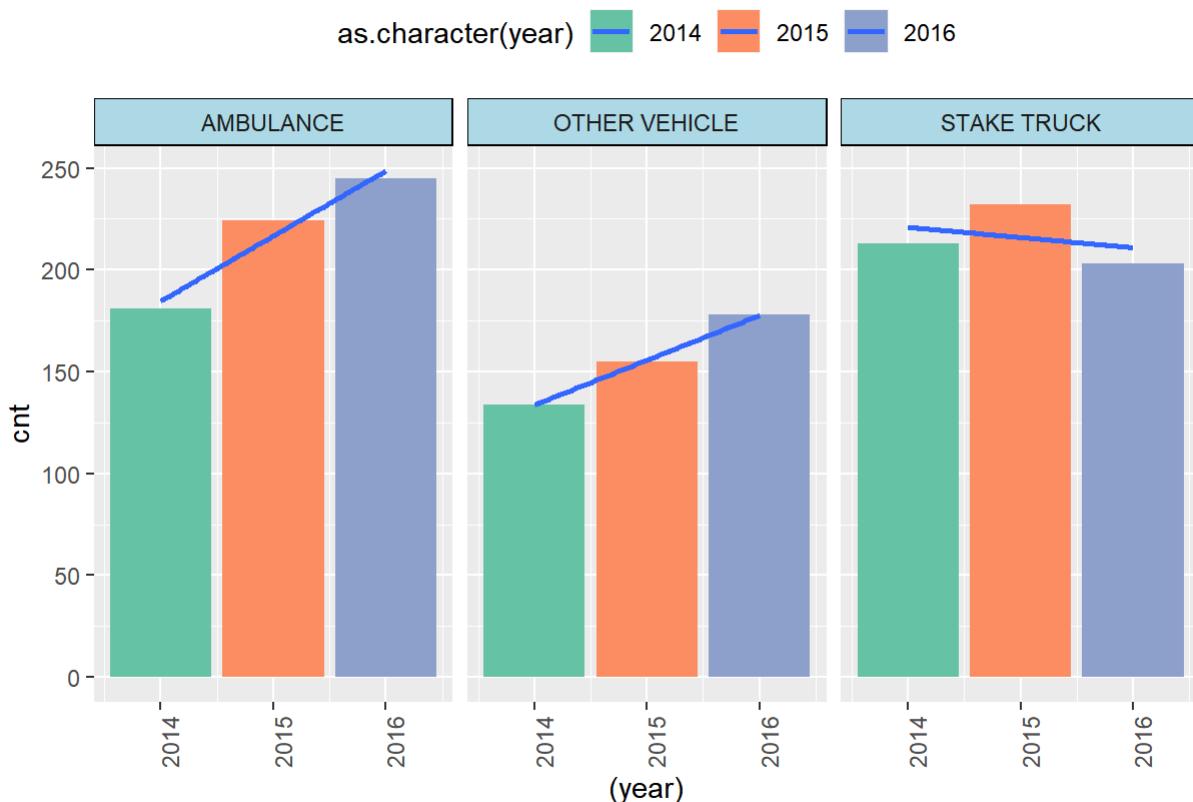
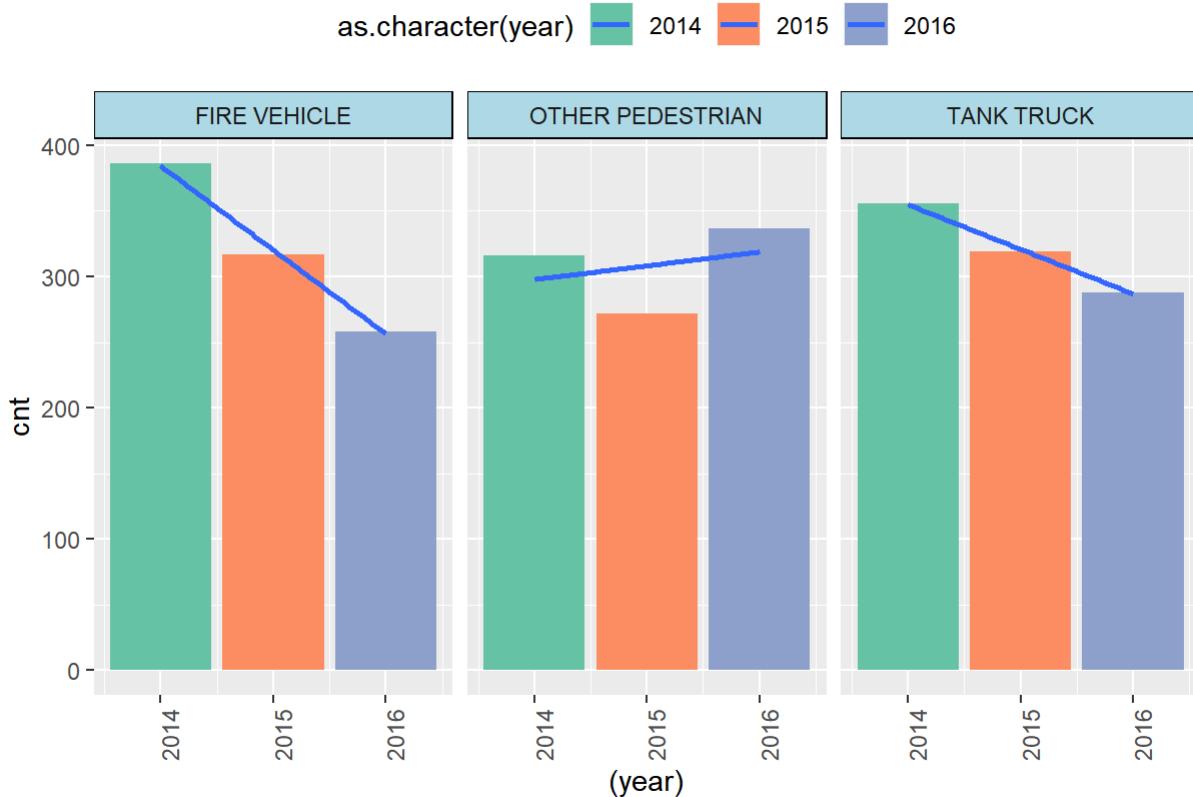
for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-(i)*9)
  print(ggplot(my_veh_dist[frm:to,],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) +geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~Vehicle.Body.Type, ncol=5, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}
```

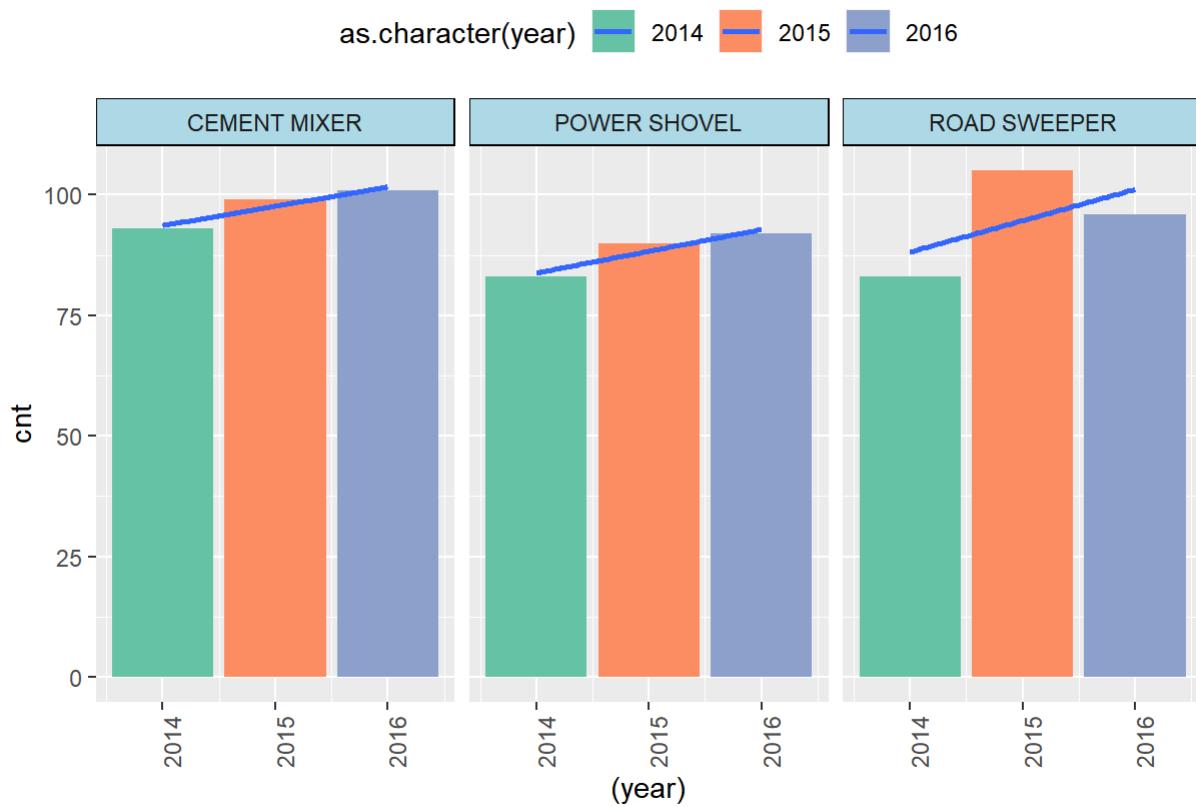







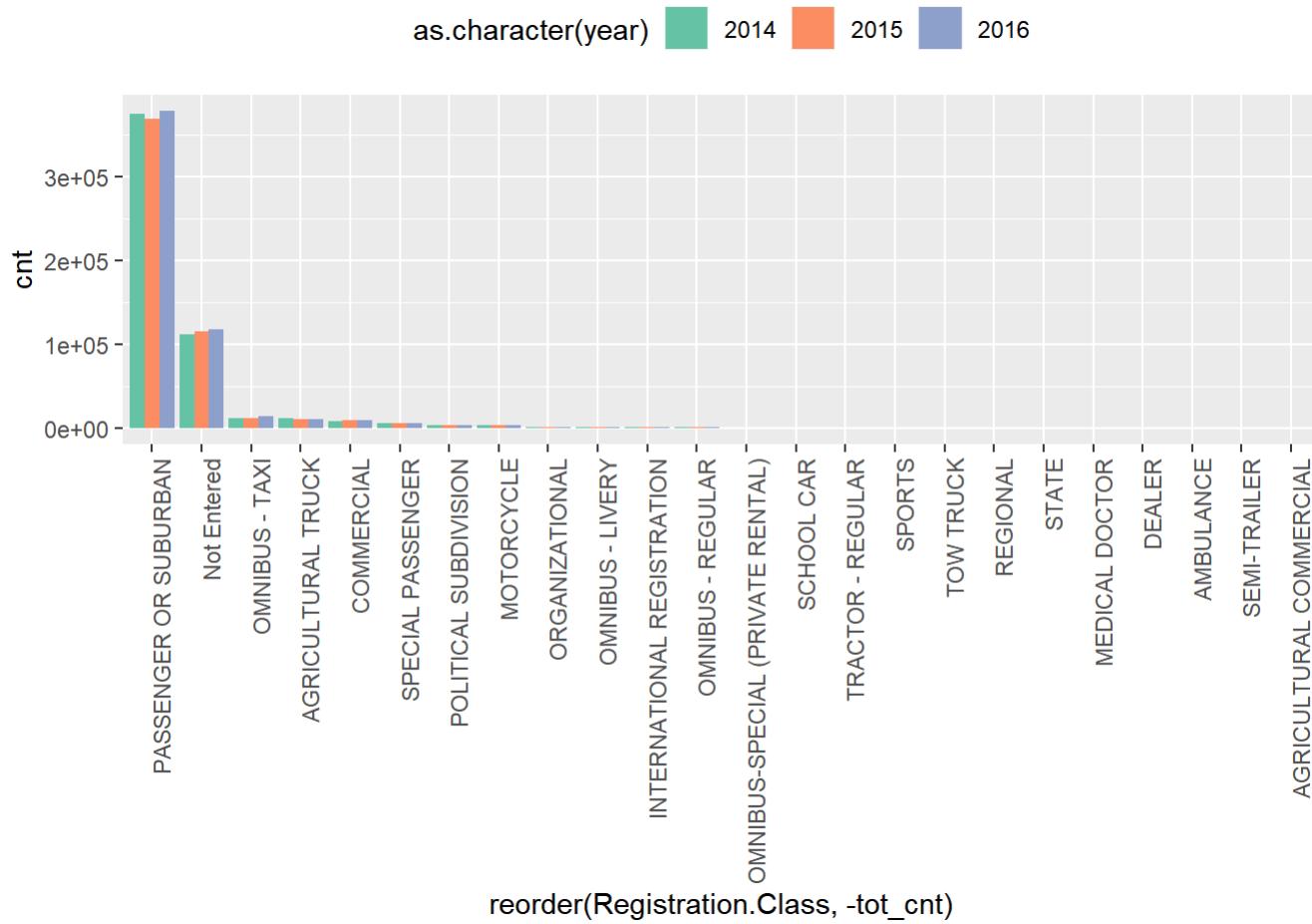






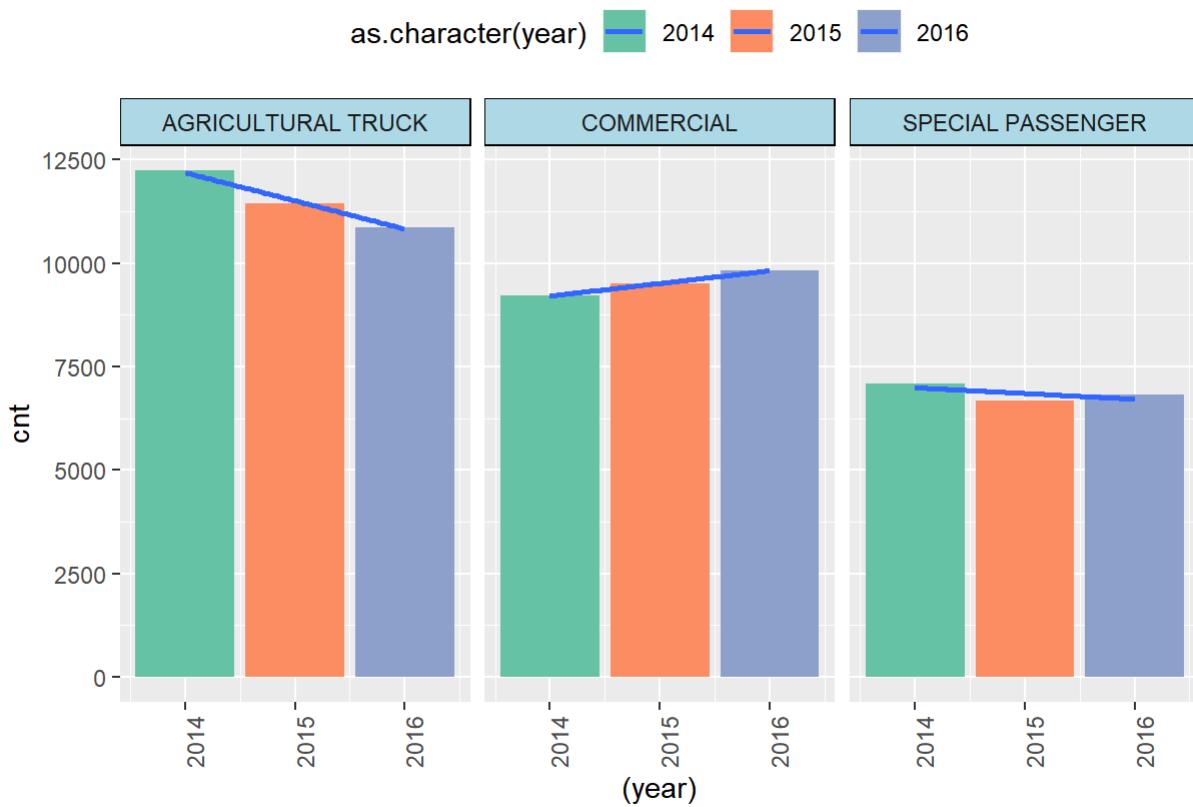
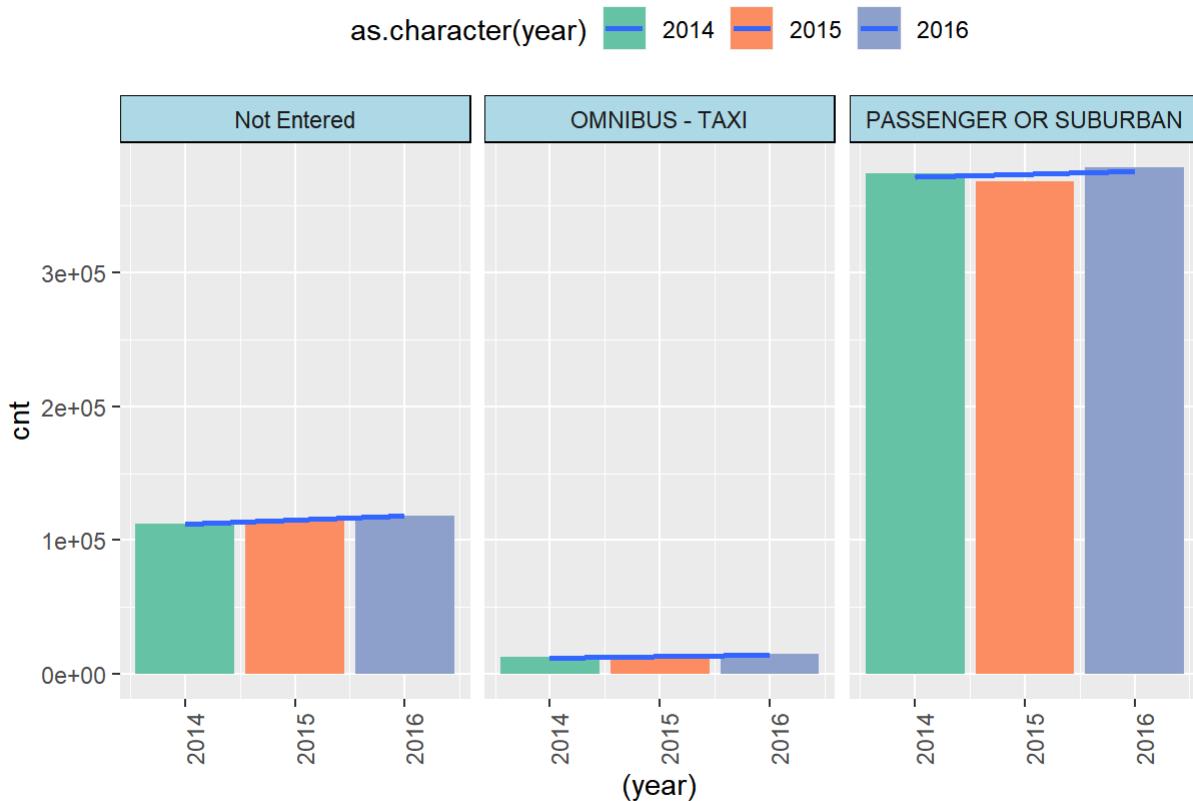
Plot Registration.Class YOY breakdown

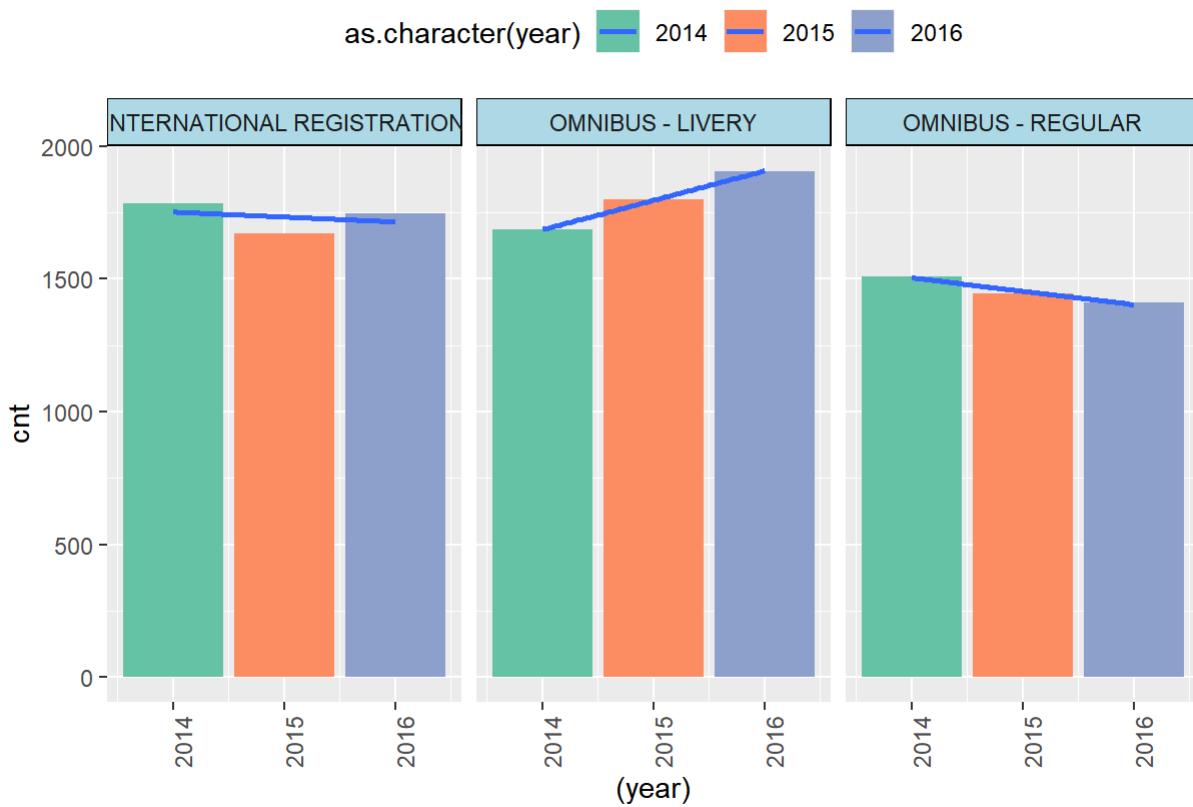
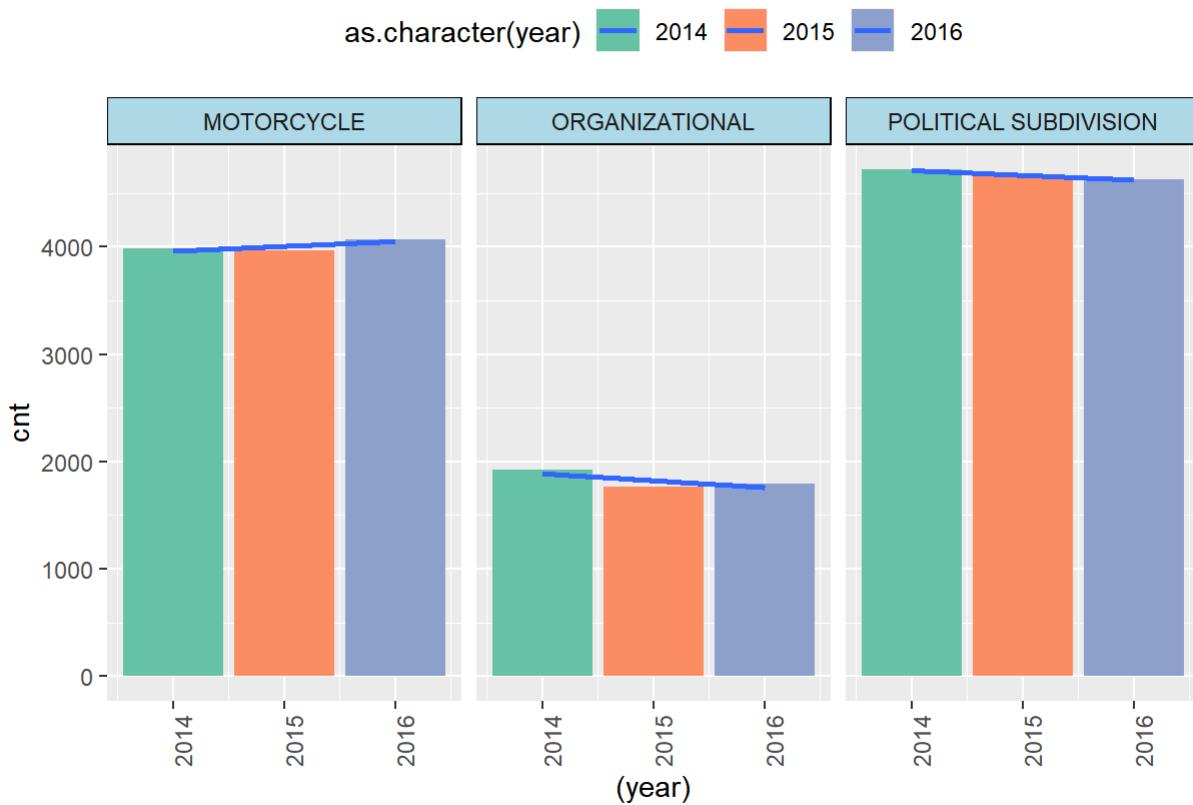
```
ggplot(my_reg_dist[which(my_reg_dist$tot_cnt>500),],aes(x=reorder(Registration.Class,-tot_cnt),
y=cnt,fill=as.character(year))) +geom_bar(stat="identity", position = position_dodge(preserve =
"total")) + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+ scale_fill_brewer(palette = "Set2")
```

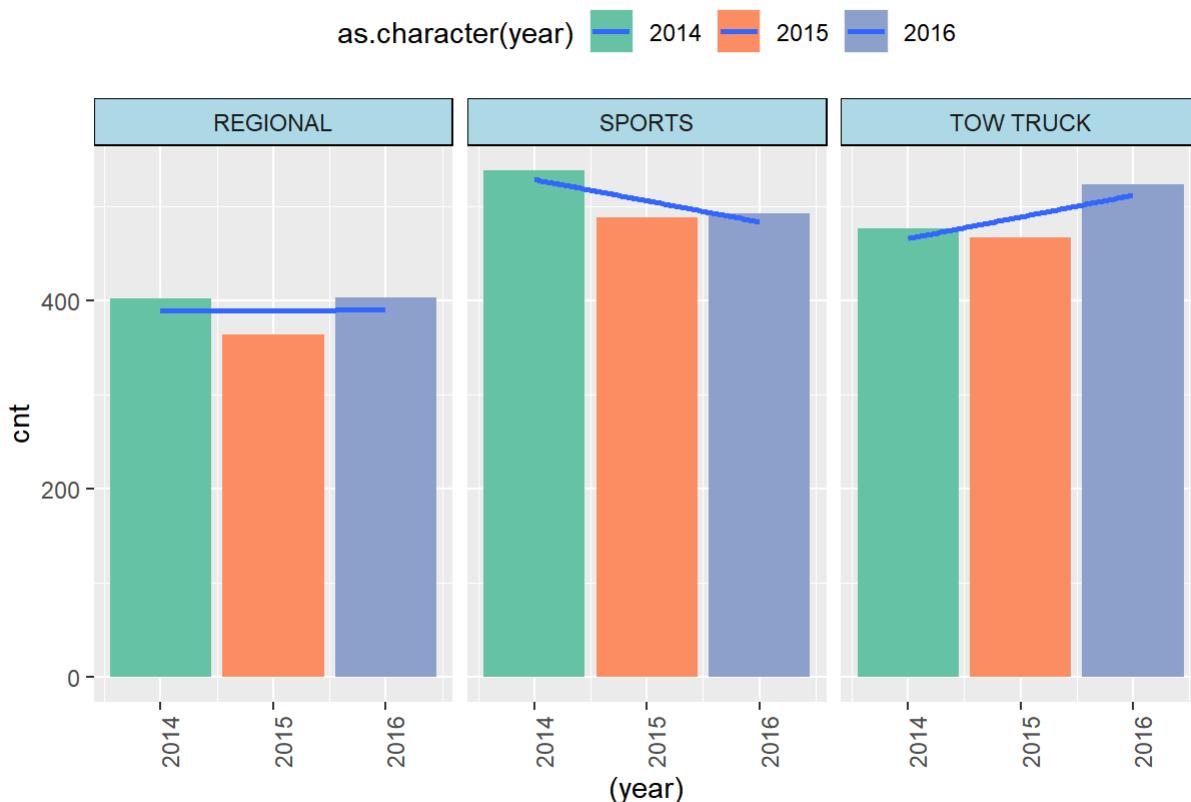
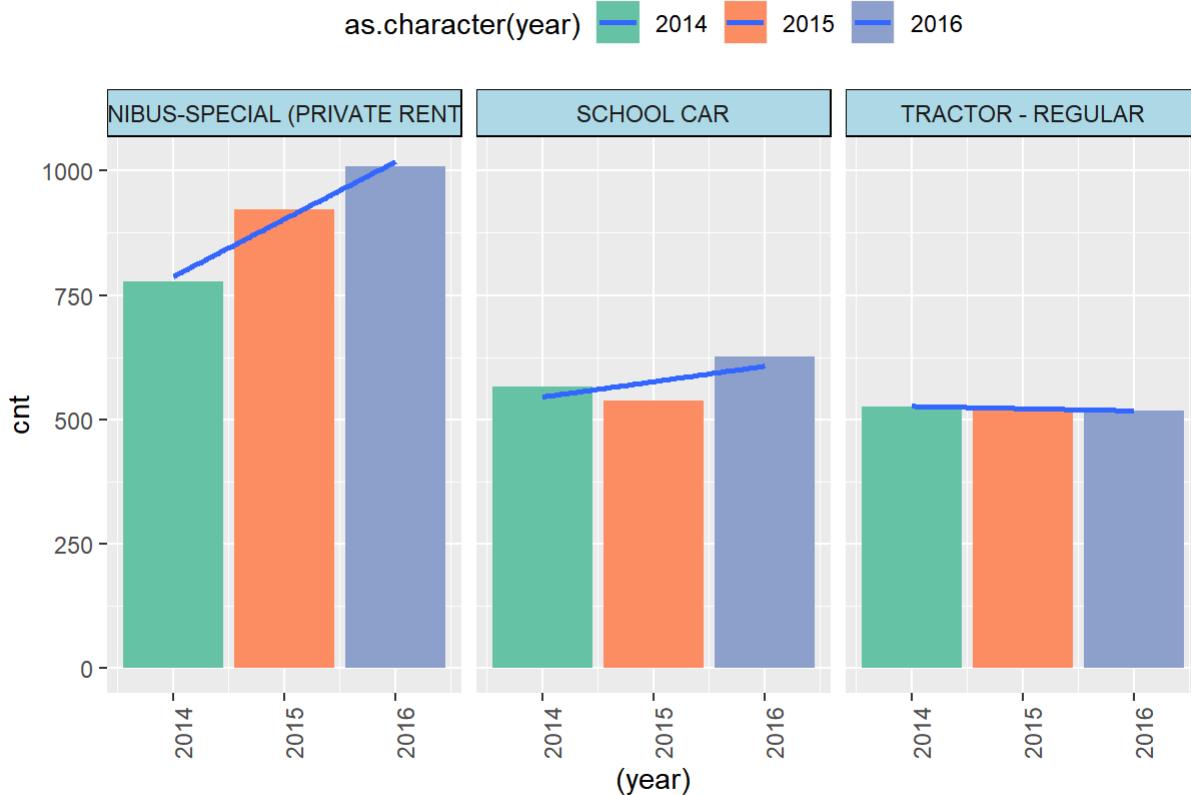


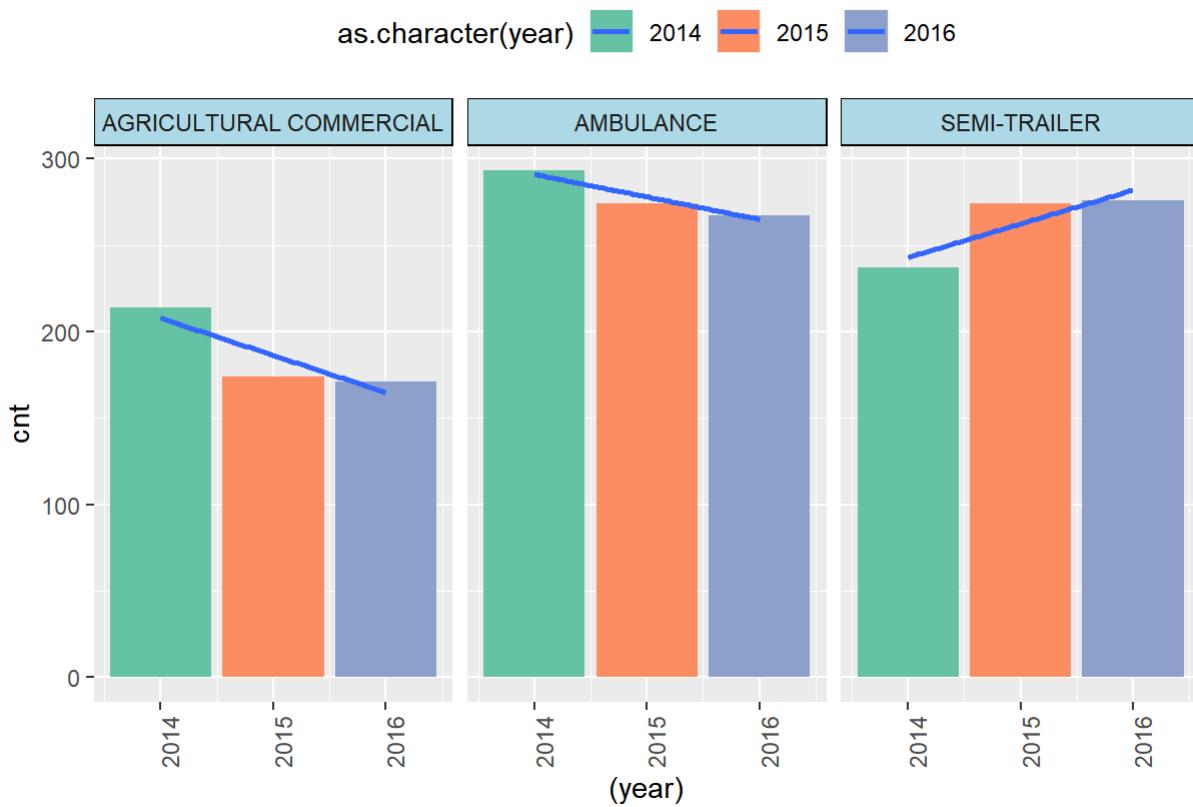
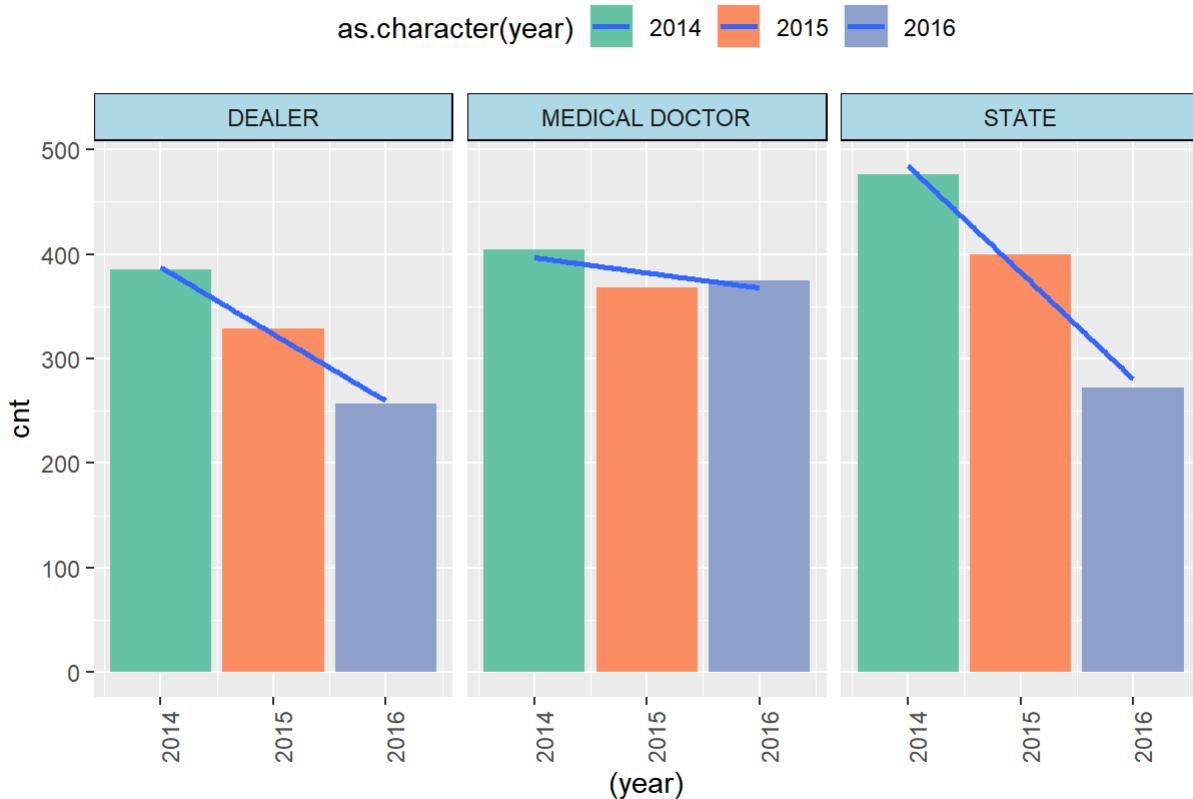
```
seq <-as.integer(nrow(my_reg_dist[which(my_reg_dist$tot_cnt>300 & my_reg_dist$yr_cnt==3),])/9)+1

for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-((i)*9)
  print(ggplot(my_reg_dist[frm:to],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) +geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~Registration.Class, ncol=5, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}
```





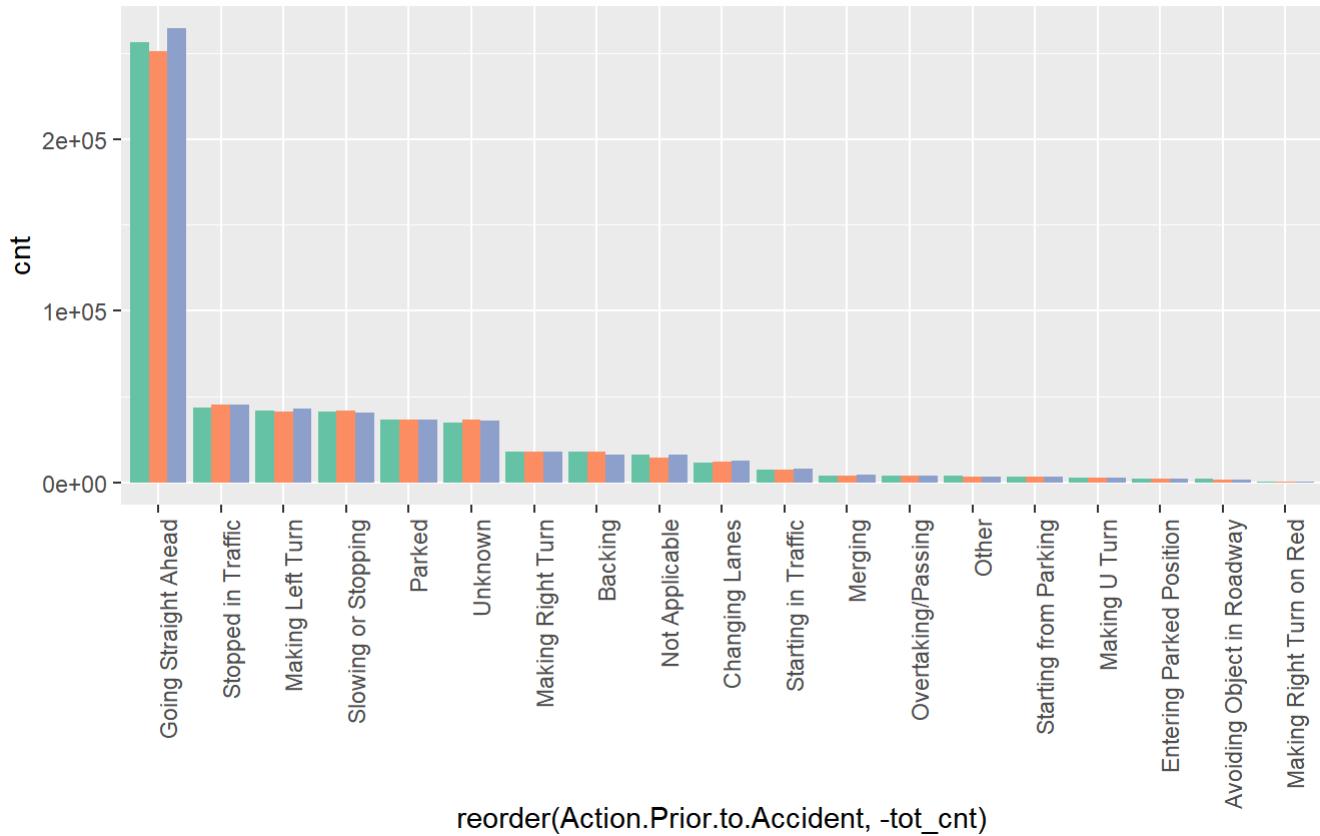




Plot Action.Prior.to.Accident YOY breakdown

```
ggplot(my_act_dist[which(my_act_dist$tot_cnt>500),],aes(x=reorder(Action.Prior.to.Accident,-tot_cnt), y=cnt,fill=as.character(year))) +geom_bar(stat="identity", position = position_dodge(preserve = "total")) + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+ scale_fill_brewer(palette = "Set2")
```

as.character(year) 2014 2015 2016

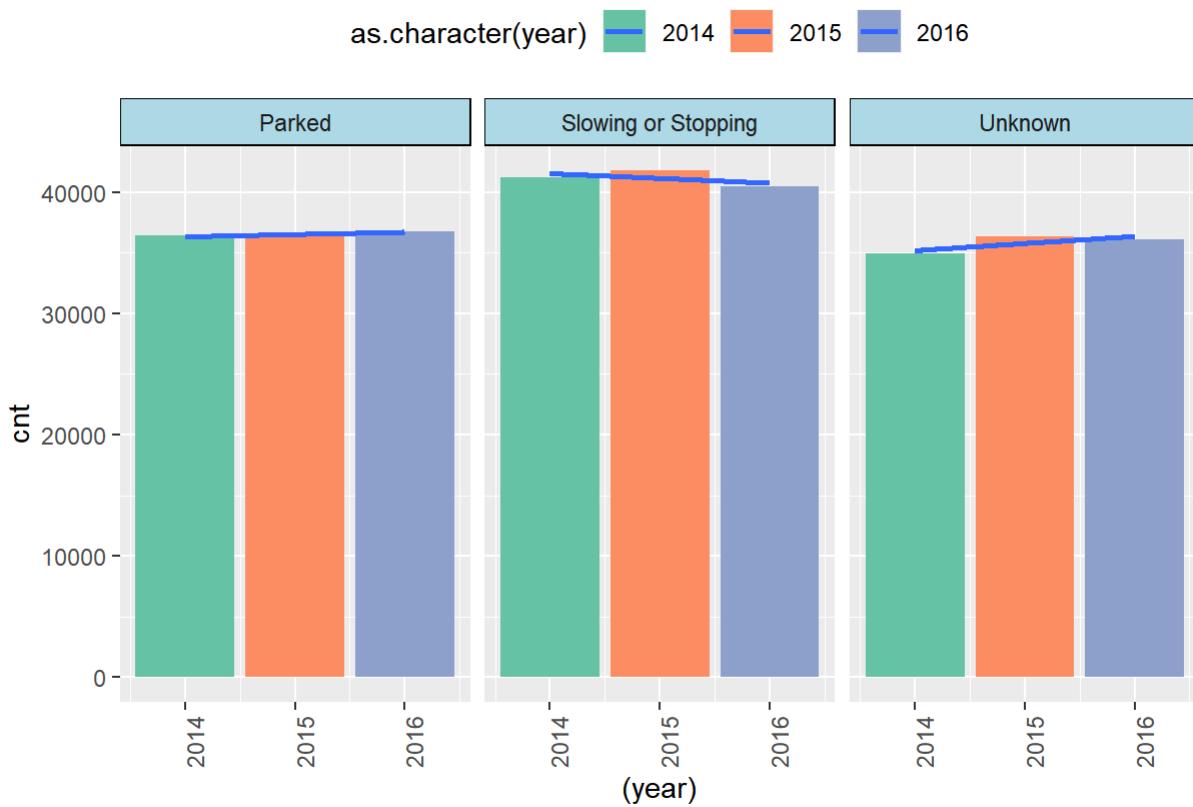
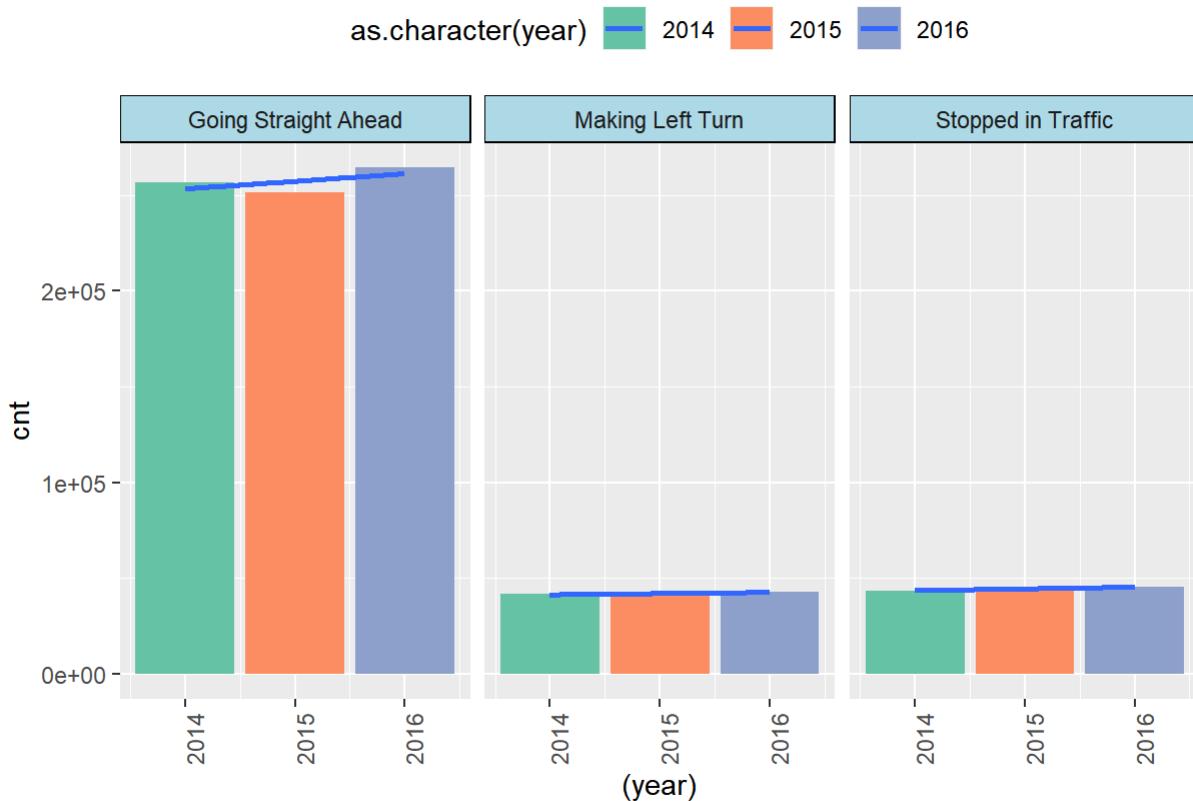


```

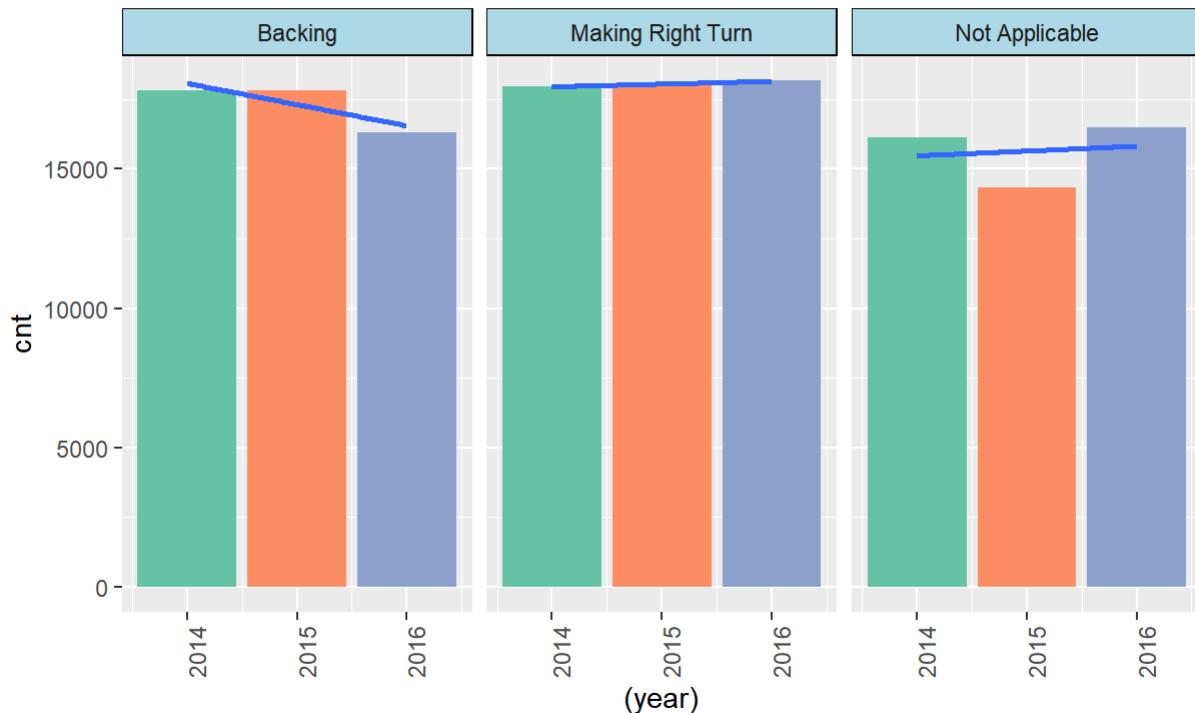
seq <-as.integer(nrow(my_act_dist)[which(my_act_dist$tot_cnt>300 & my_act_dist$yr_cnt==3),])/9)+1

for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-((i)*9)
print(ggplot(my_act_dist[frm:to,],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) +geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~Action.Prior.to.Accident, ncol=5, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}

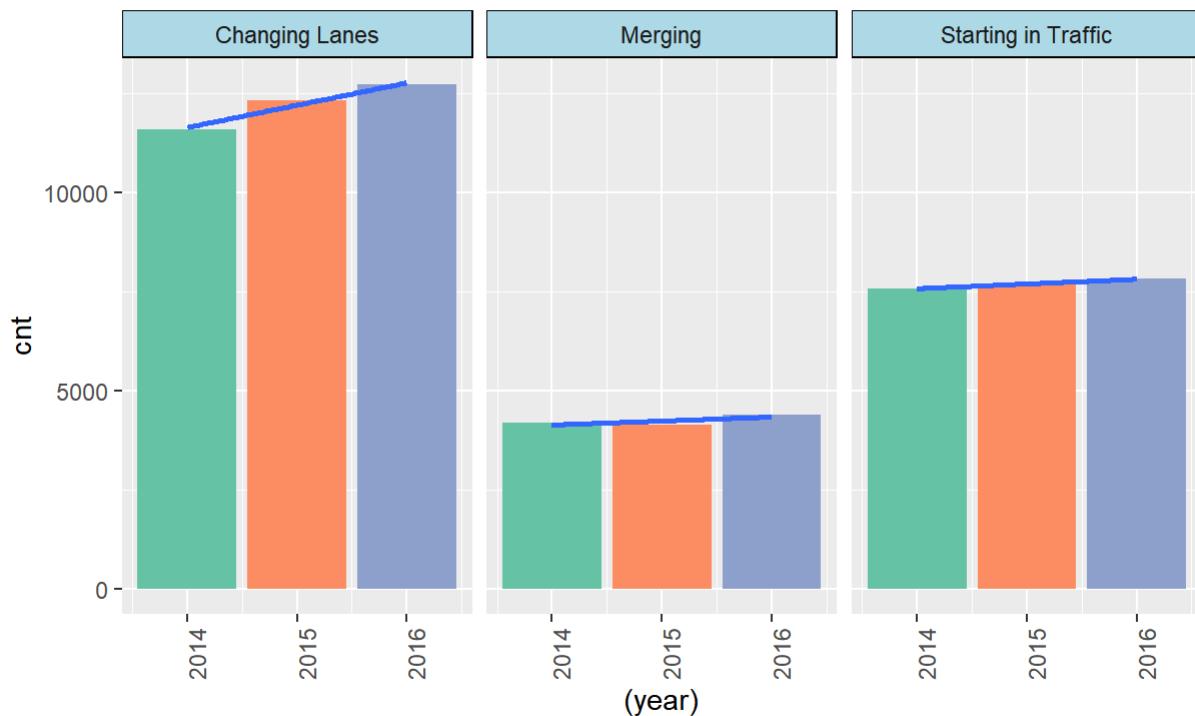
```

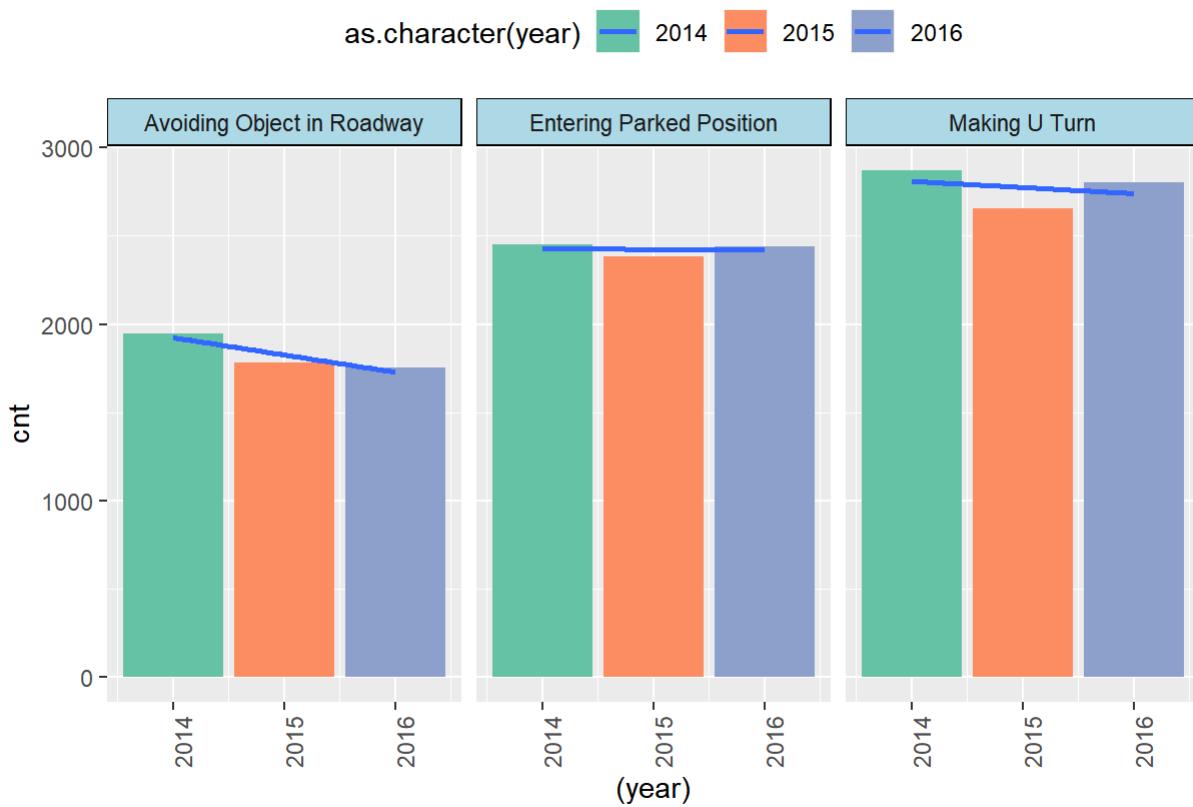
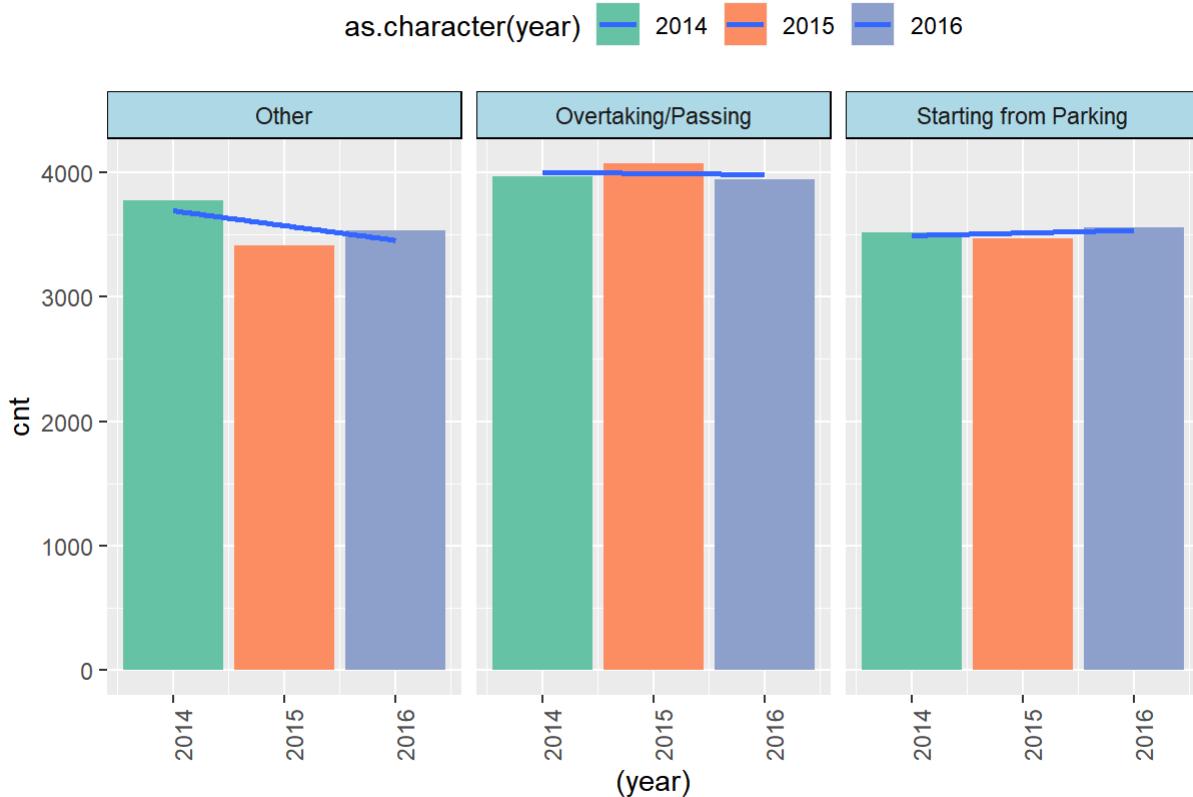



as.character(year) 2014 2015 2016



as.character(year) 2014 2015 2016

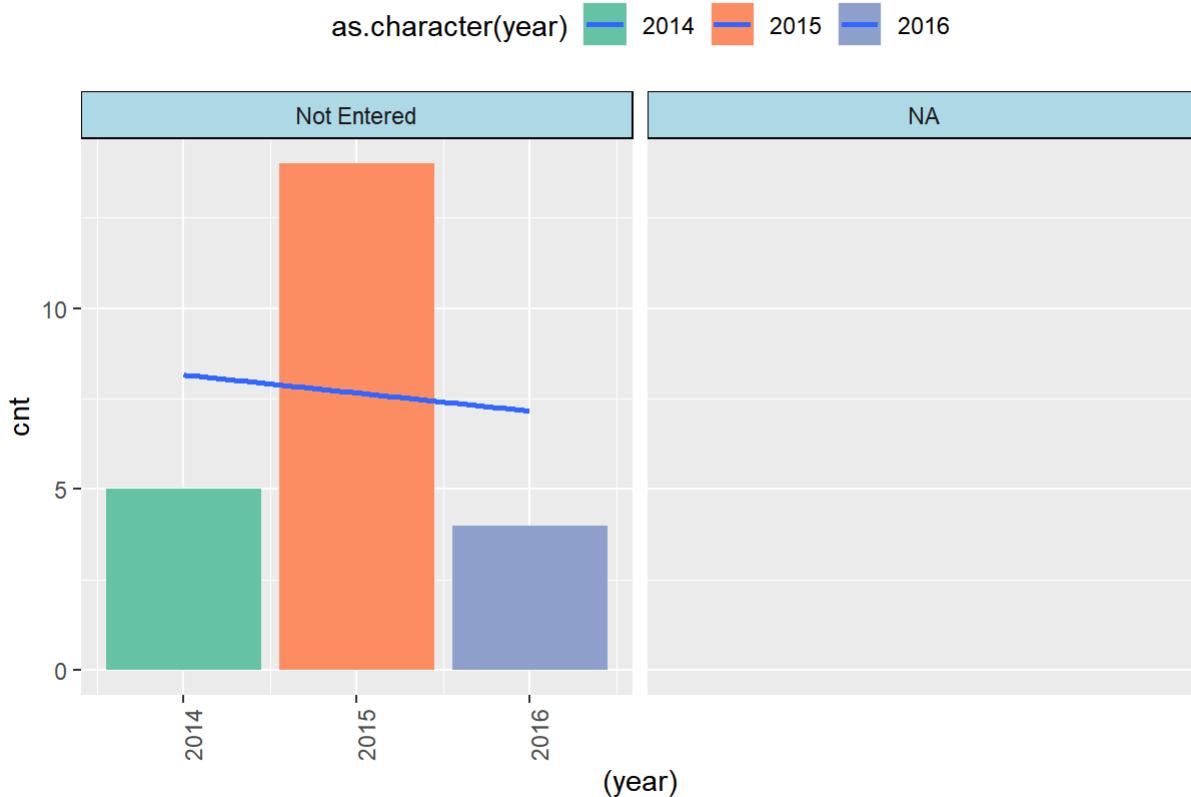






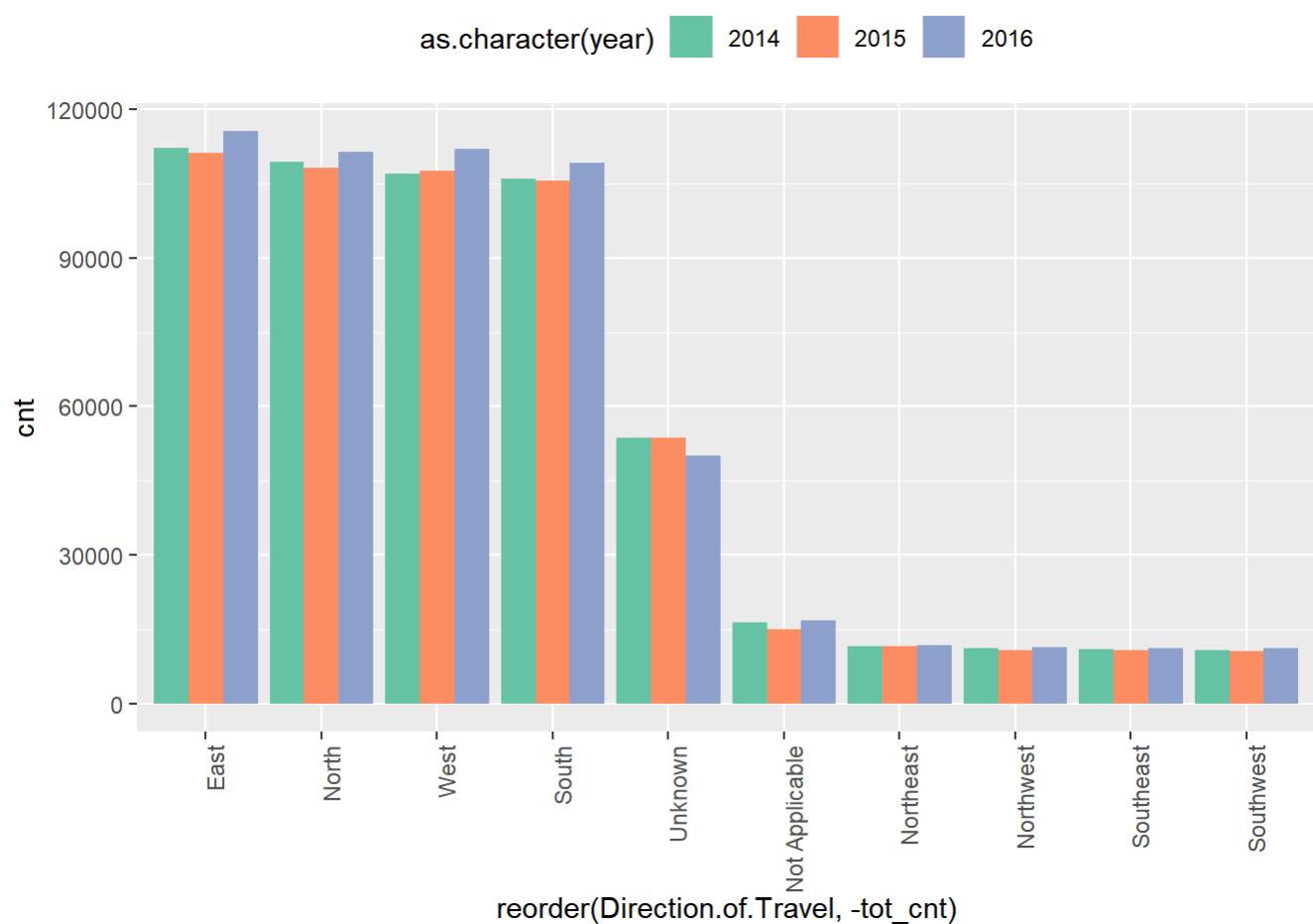
```
## Warning: Removed 6 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6 rows containing missing values (position_stack).
```



Plot Direction.of.Travel YOY breakdown

```
ggplot(my_dir_dist[which(my_dir_dist$tot_cnt>500),],aes(x=reorder(Direction.of.Travel,-tot_cnt),  
y=cnt,fill=as.character(year)))+geom_bar(stat="identity", position = position_dodge(preserve  
="total"))+ theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+  
scale_fill_brewer(palette = "Set2")
```

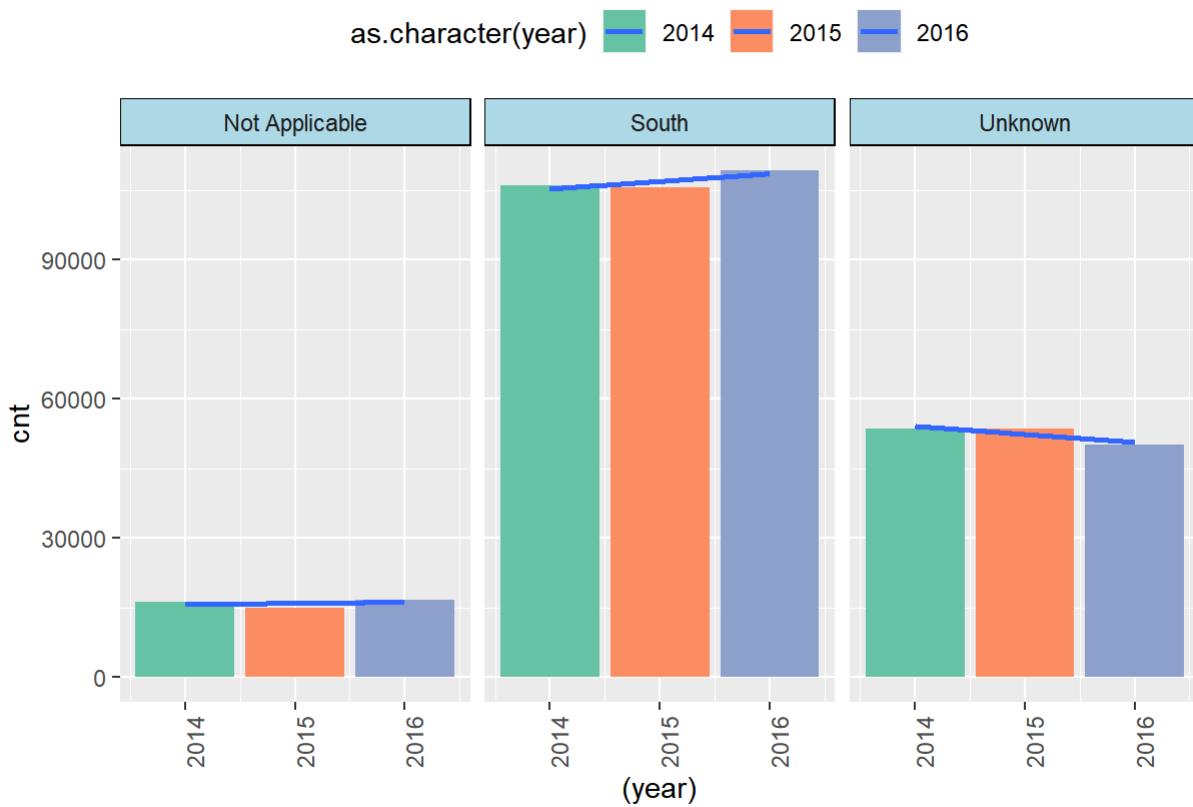
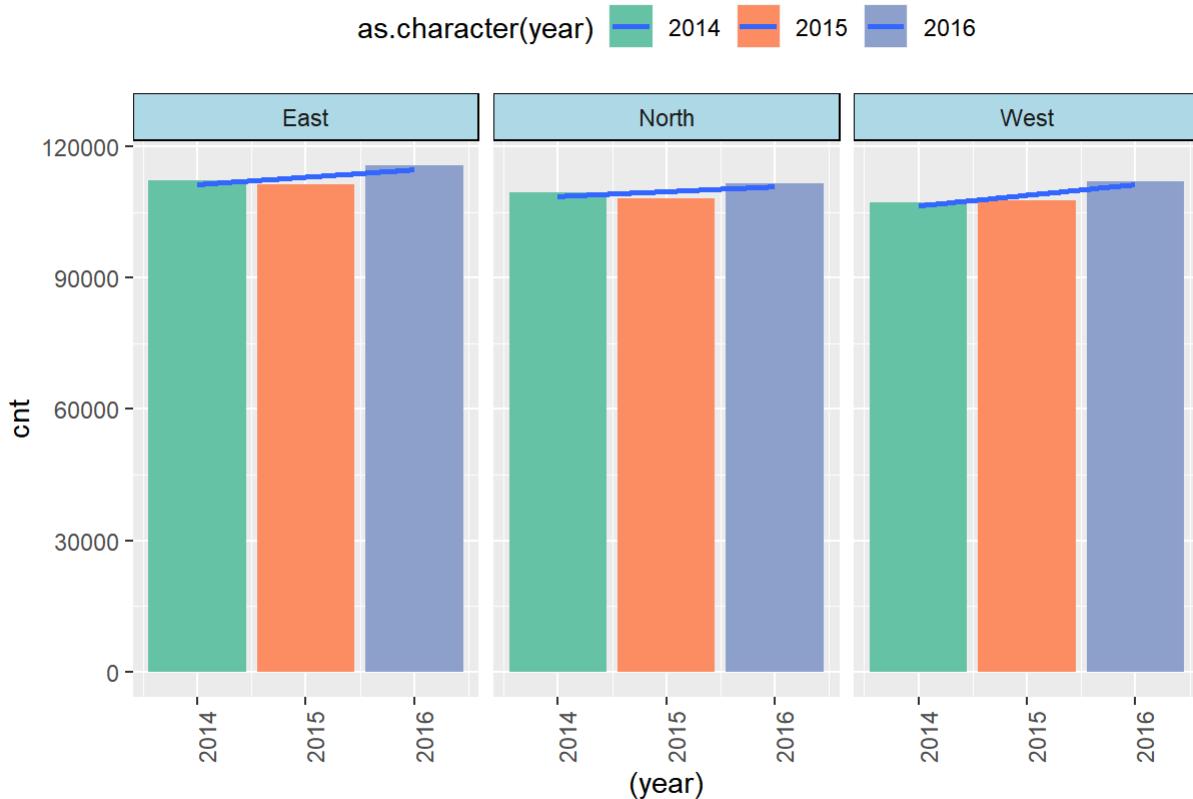


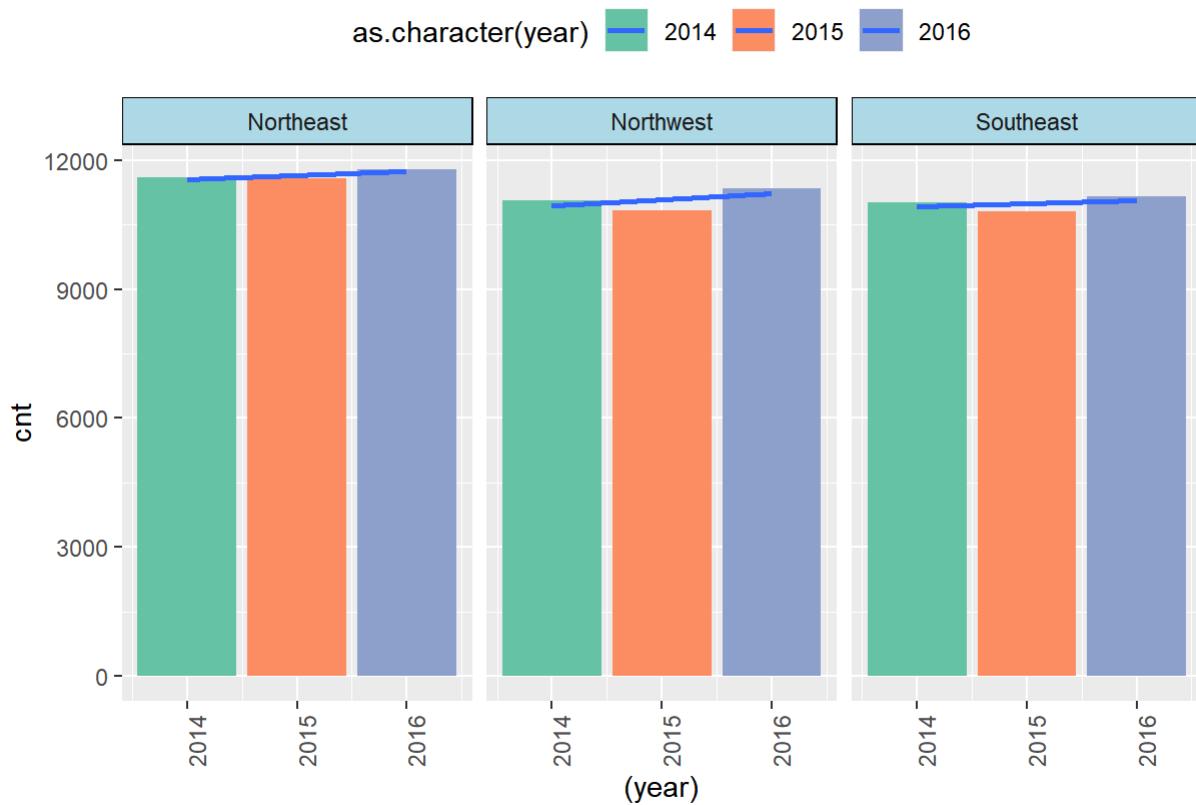
```

seq <-as.integer(nrow(my_dir_dist)[which(my_dir_dist$tot_cnt>300 & my_dir_dist$yr_cnt==3),])/9)+1

for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-((i)*9)
  print(ggplot(my_dir_dist[frm:to,],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) + geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~Direction.of.Travel, ncol=5, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}

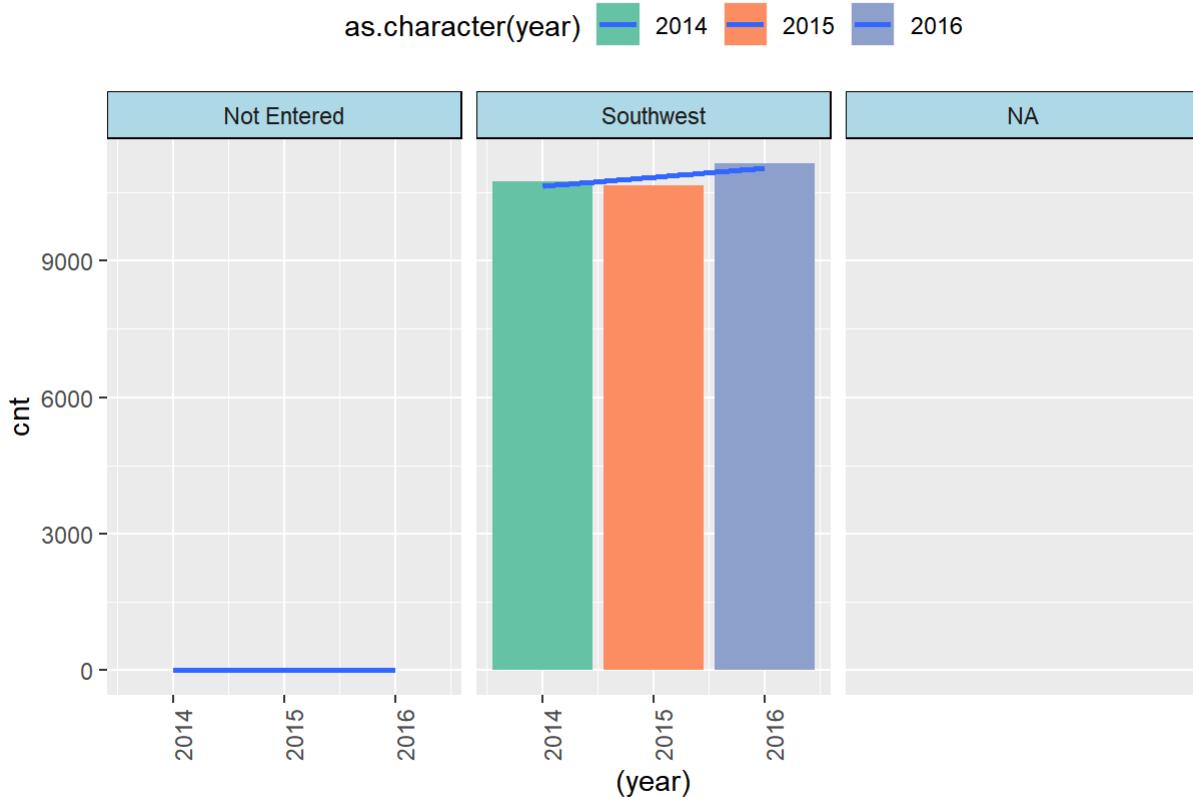
```



```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

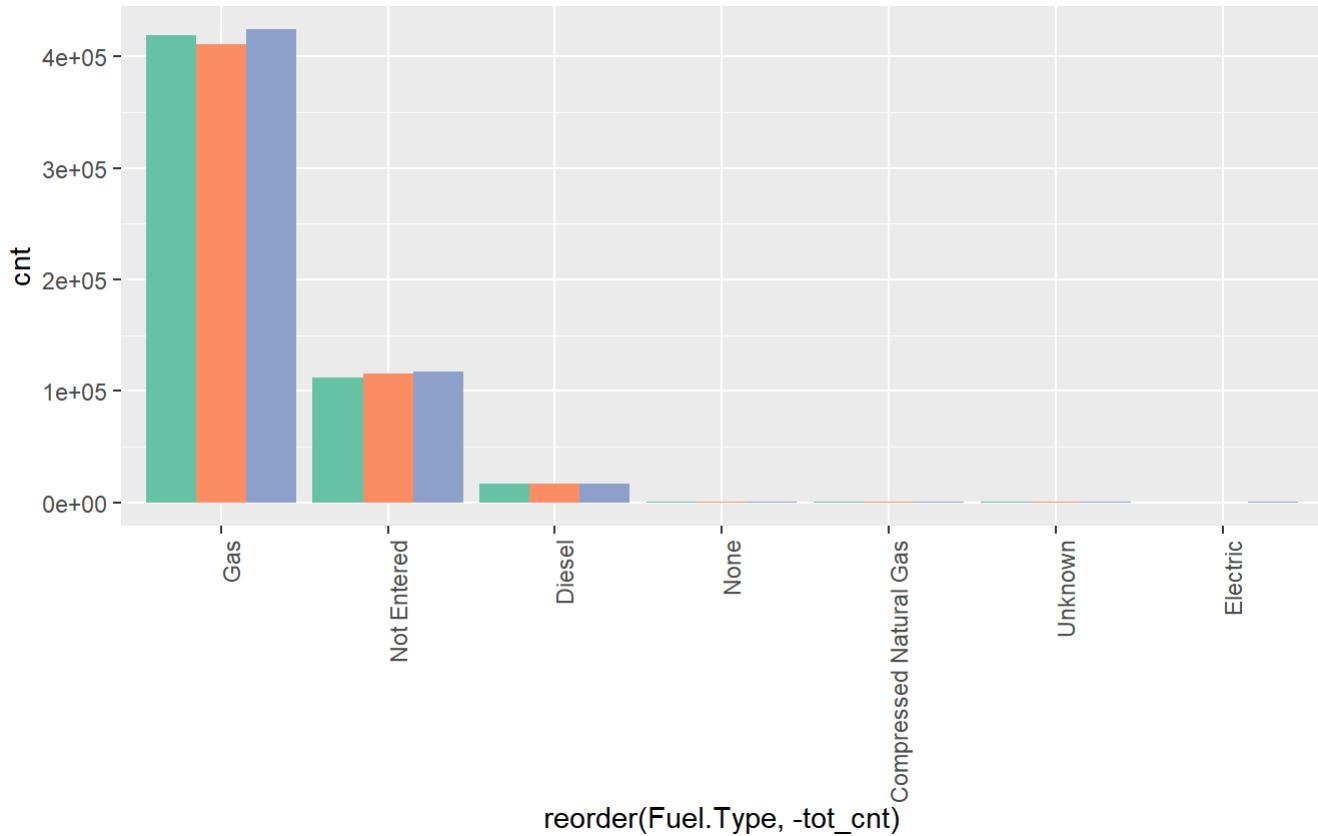
```
## Warning: Removed 3 rows containing missing values (position_stack).
```



Plot Fuel.Type YOY breakdown

```
ggplot(my_fuel_dist[which(my_fuel_dist$tot_cnt>500),],aes(x=reorder(Fuel.Type,-tot_cnt), y=cnt,f  
ill=as.character(year)) +geom_bar(stat="identity", position = position_dodge(preserve = "total"  
)) + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+ scale_fi  
ll_brewer(palette = "Set2")
```

as.character(year) 2014 2015 2016

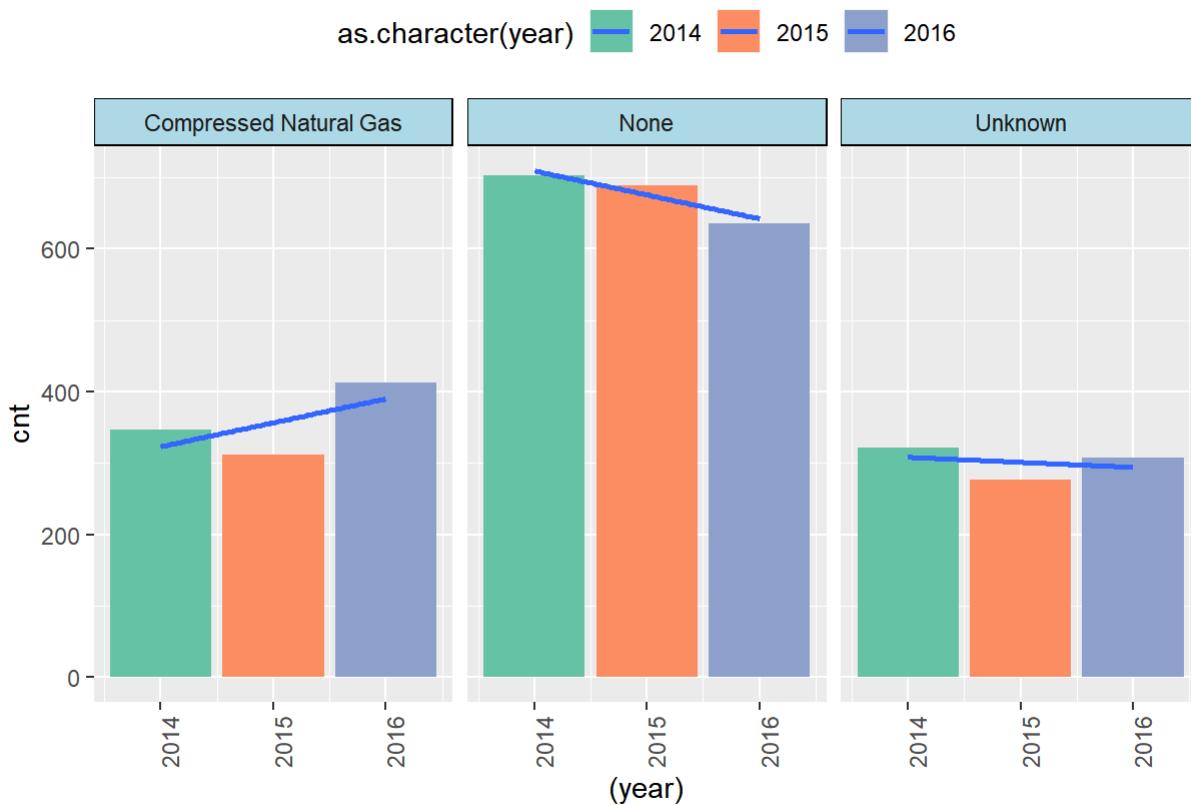
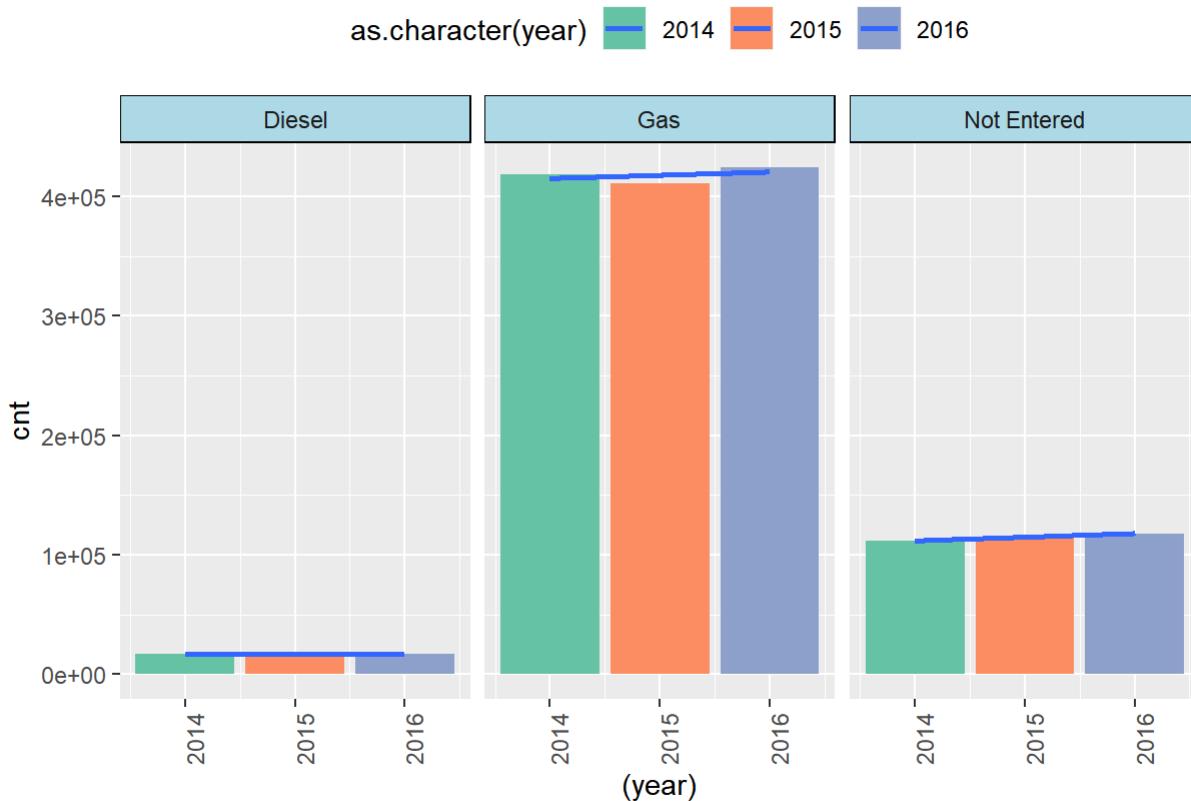


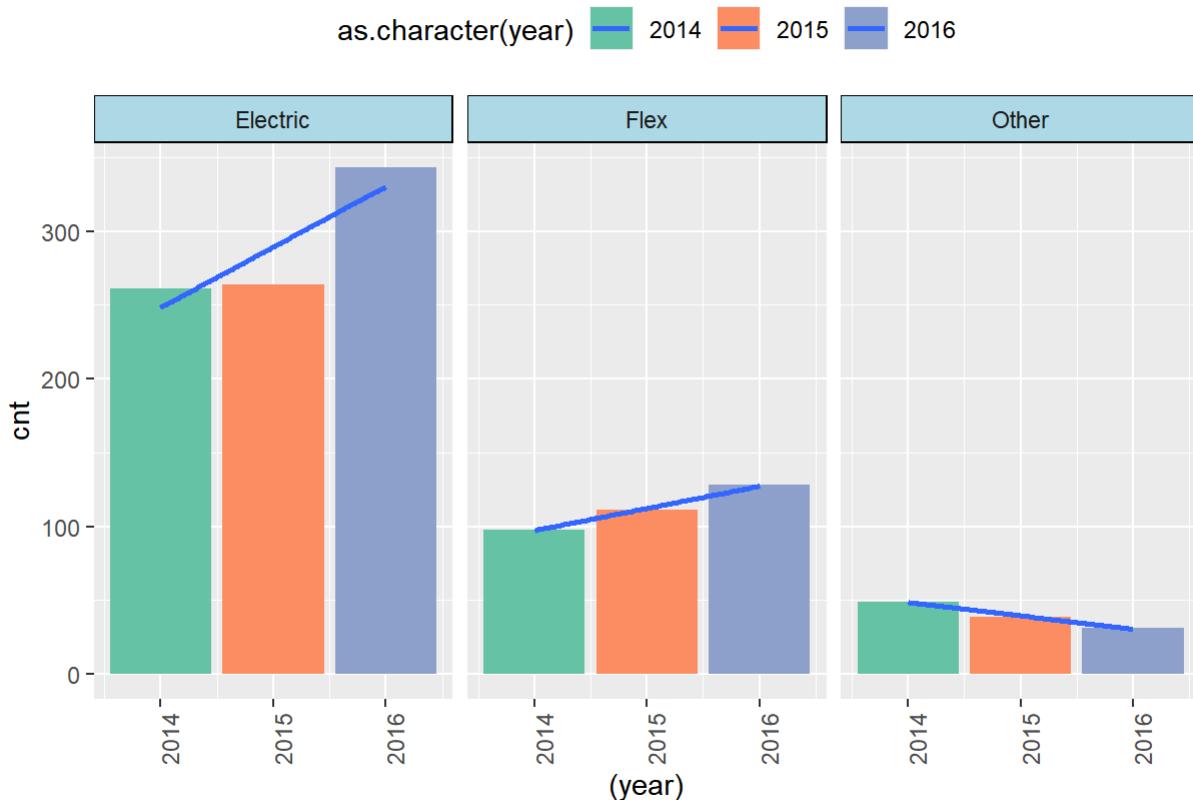
```

seq <- as.integer(nrow(my_fuel_dist[which(my_fuel_dist$tot_cnt>300 & my_fuel_dist$yr_cnt==3),])/9
)+1

for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-((i)*9)
print(ggplot(my_fuel_dist[frm:to,],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) +geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90,
  hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~Fuel.Type, ncol=5
, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}

```

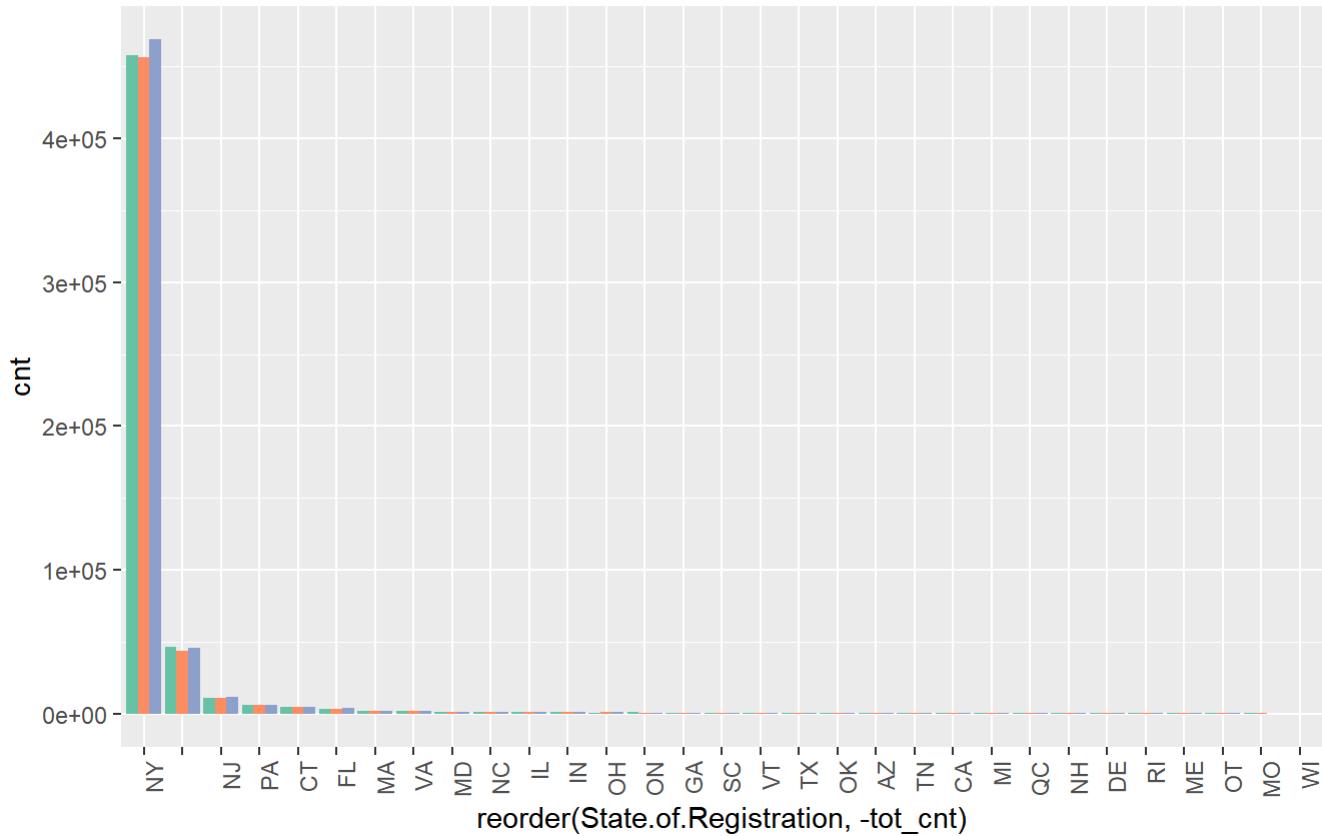





Plot State.of.Registration YOY breakdown

```
ggplot(my_st_dist[which(my_st_dist$tot_cnt>500),],aes(x=reorder(State.of.Registration,-tot_cnt),
y=cnt,fill=as.character(year))) +geom_bar(stat="identity", position = position_dodge(preserve =
"total")) + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+scale_fill_brewer(palette = "Set2")
```

as.character(year) 2014 2015 2016

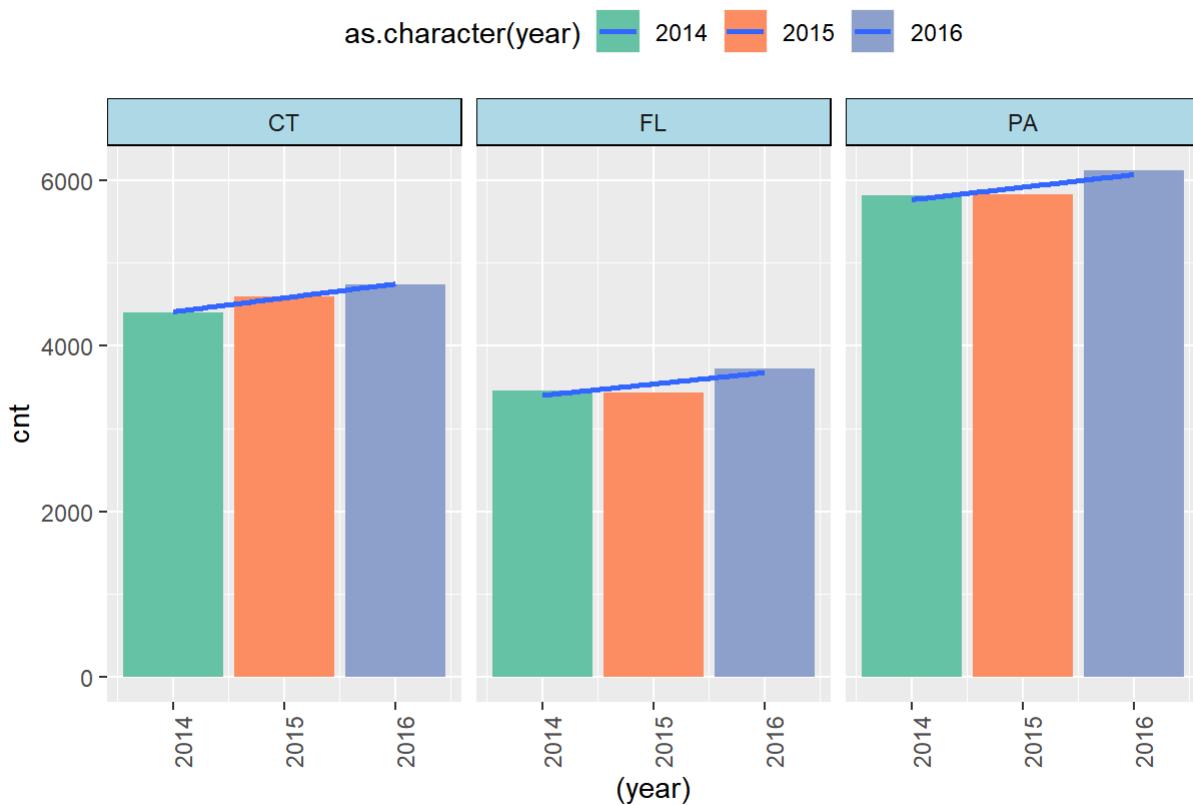
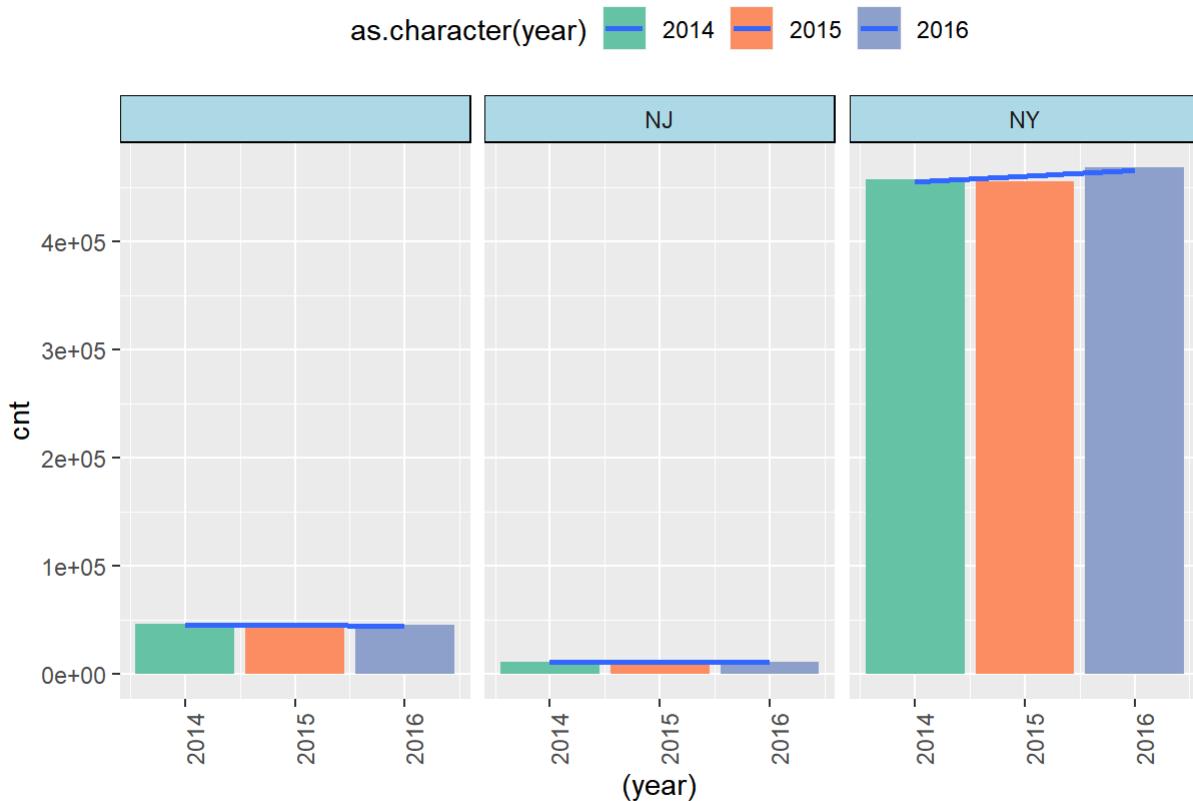


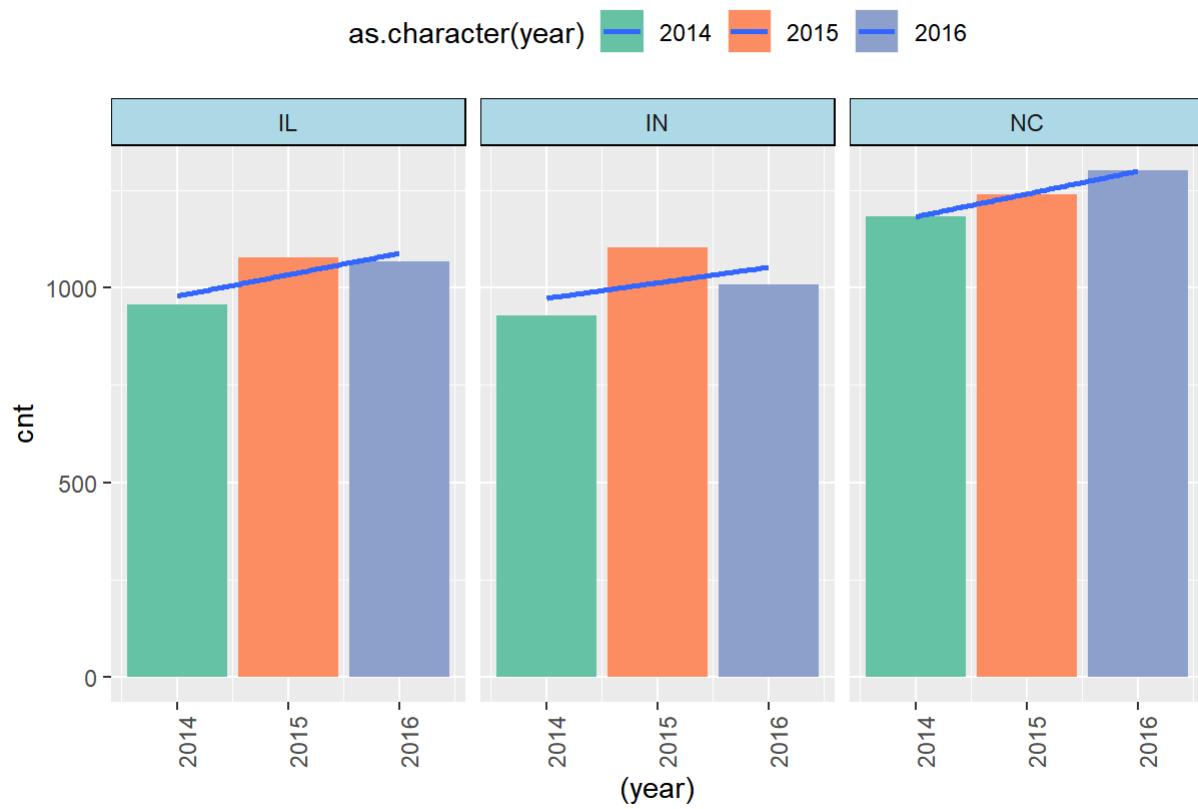
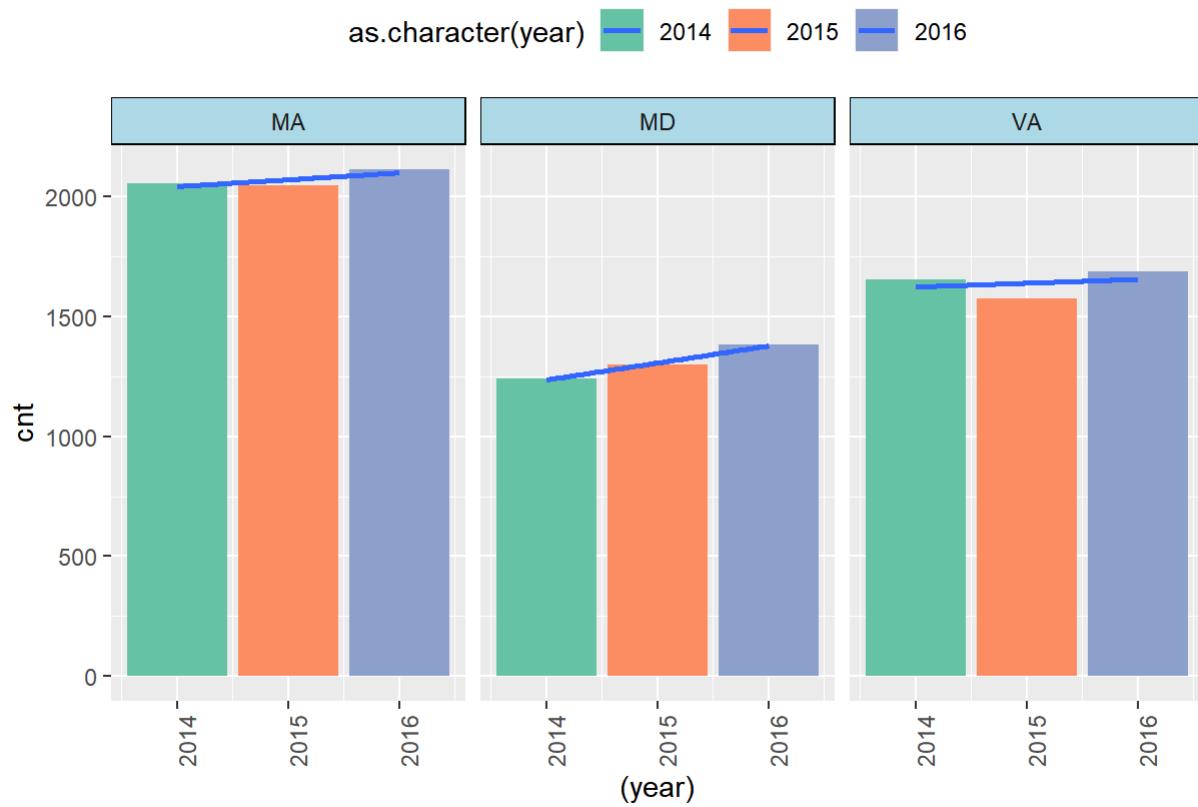
```

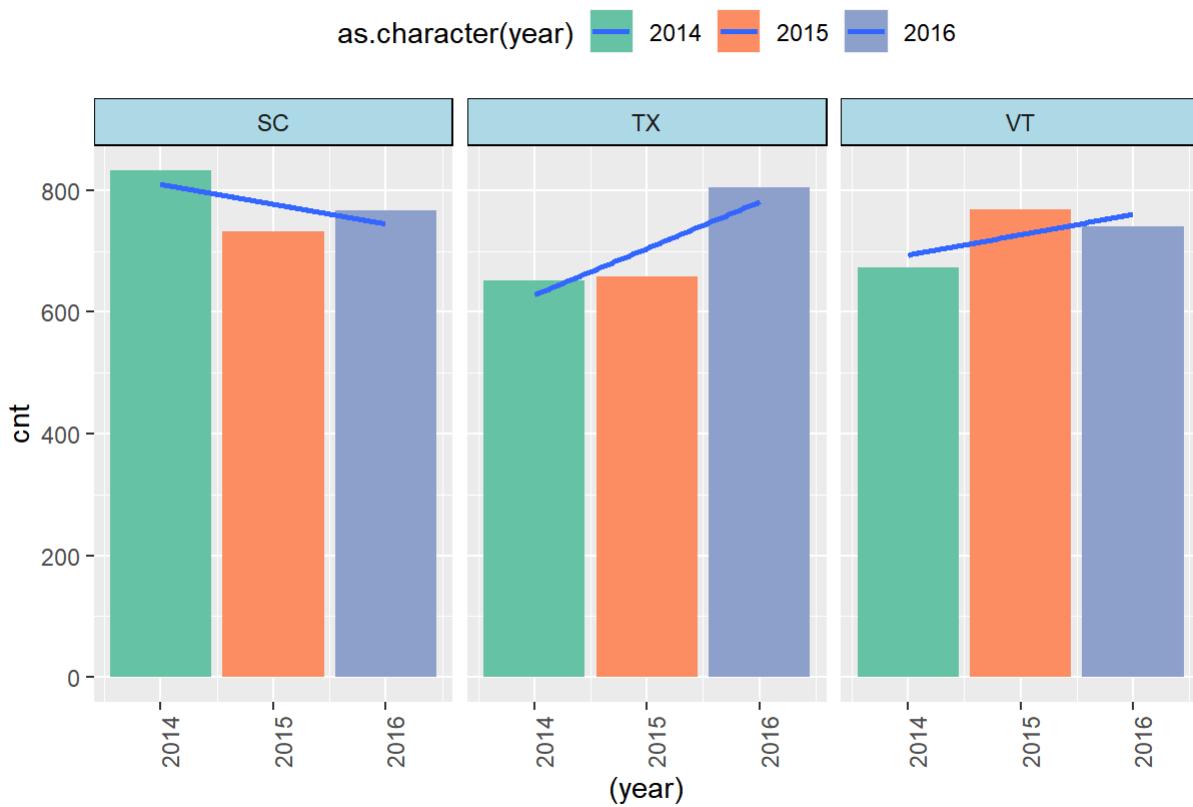
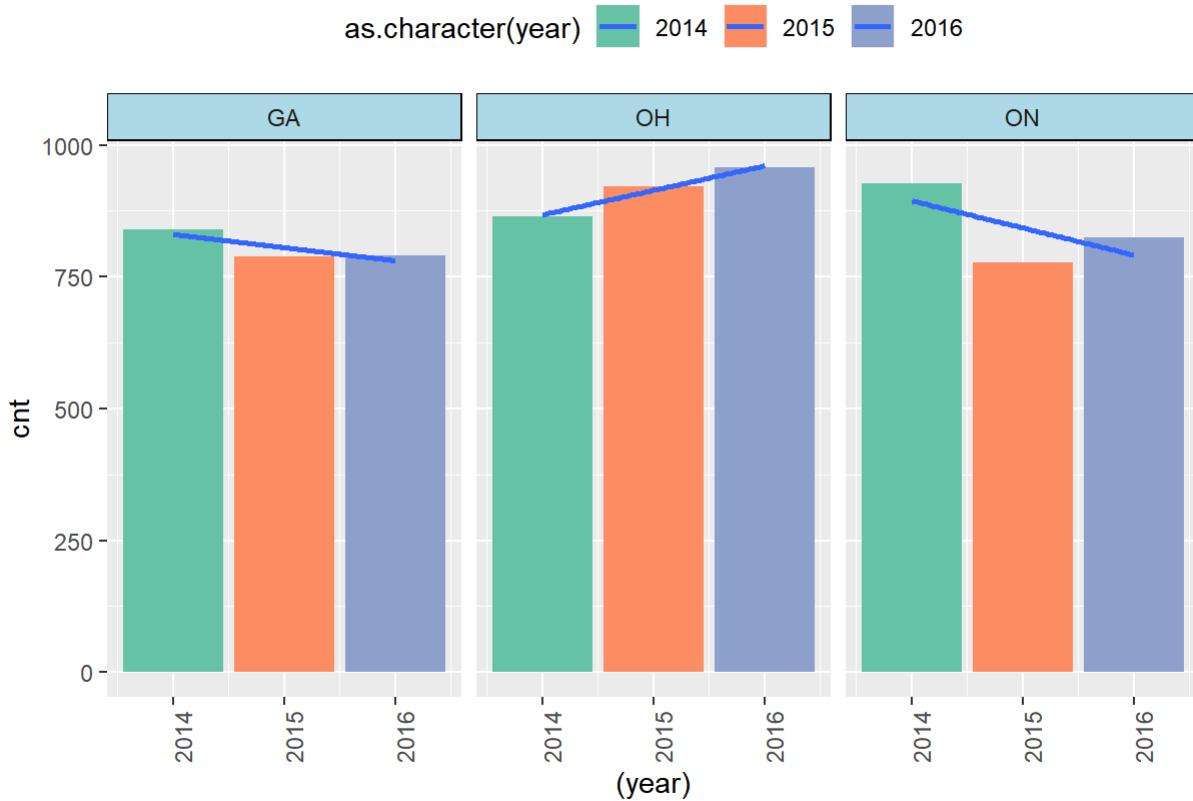
seq <-as.integer(nrow(my_st_dist)[which(my_st_dist$tot_cnt>300 & my_st_dist$yr_cnt==3),])/9)+1

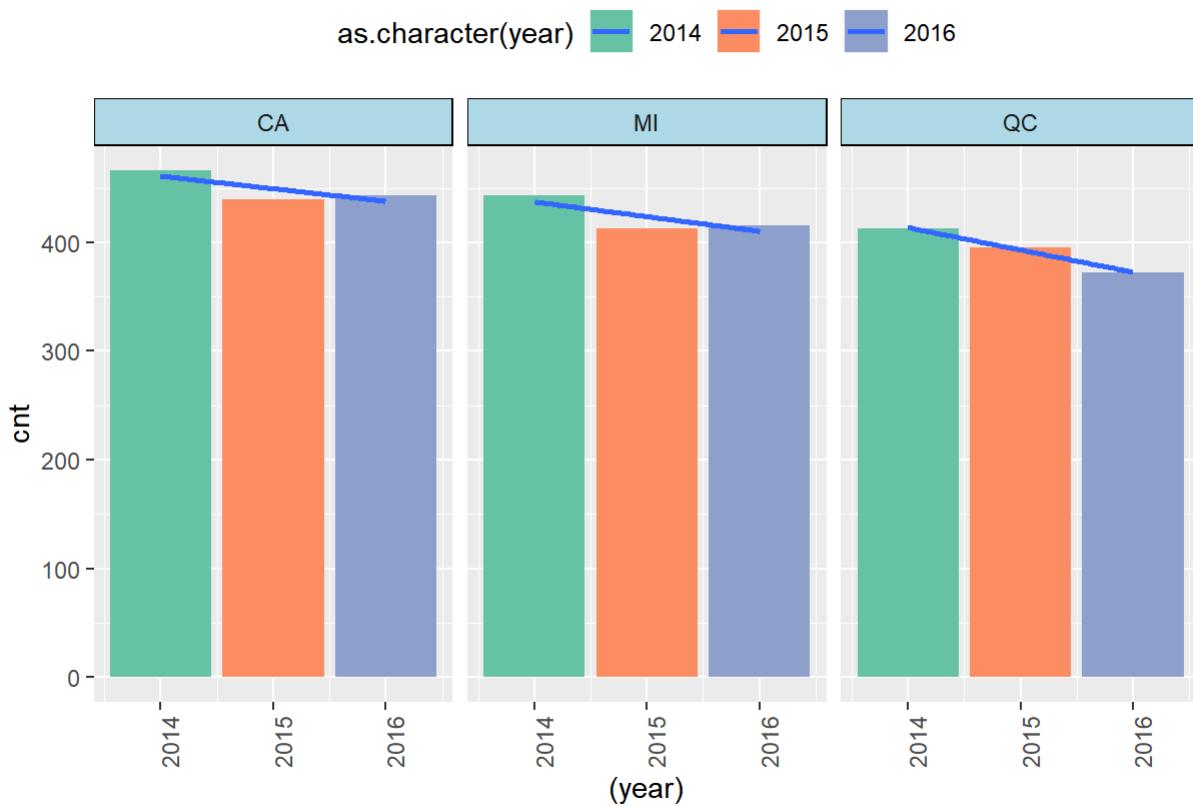
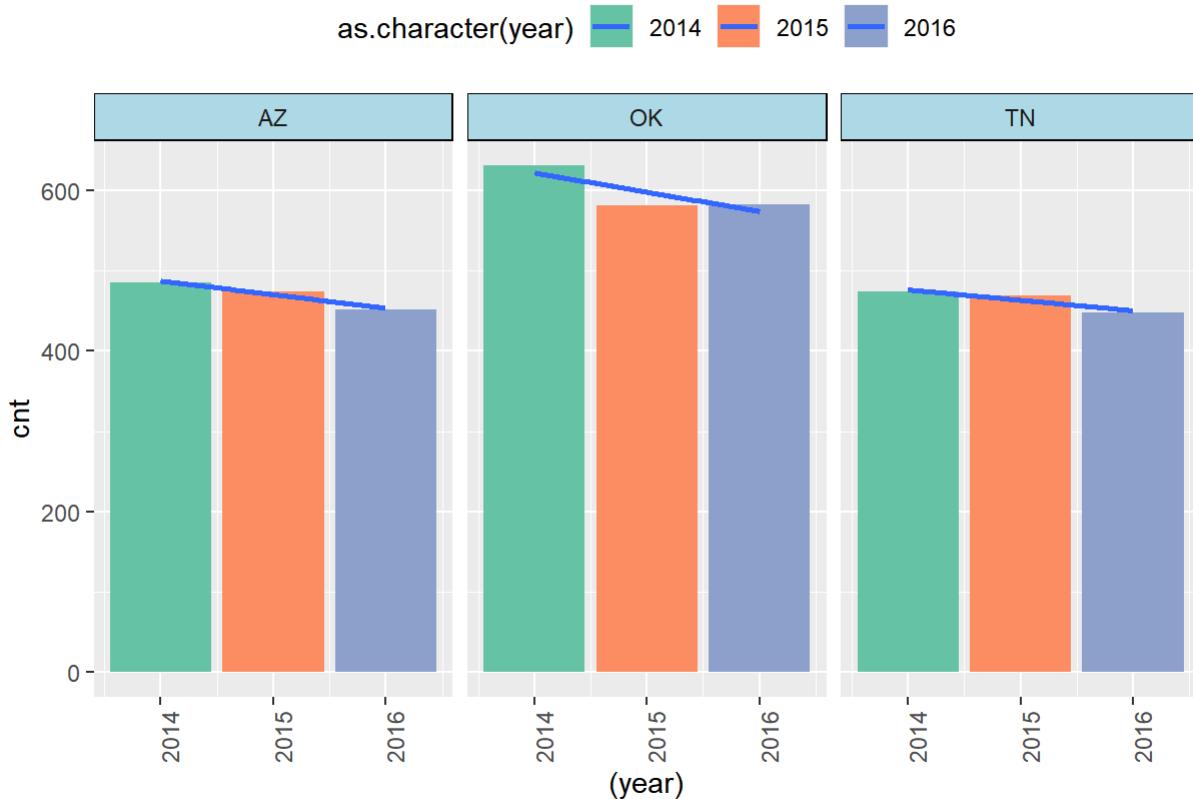
for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-((i)*9)
print(ggplot(my_st_dist[frm:to,],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) +geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~State.of.Registration, ncol=5, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}

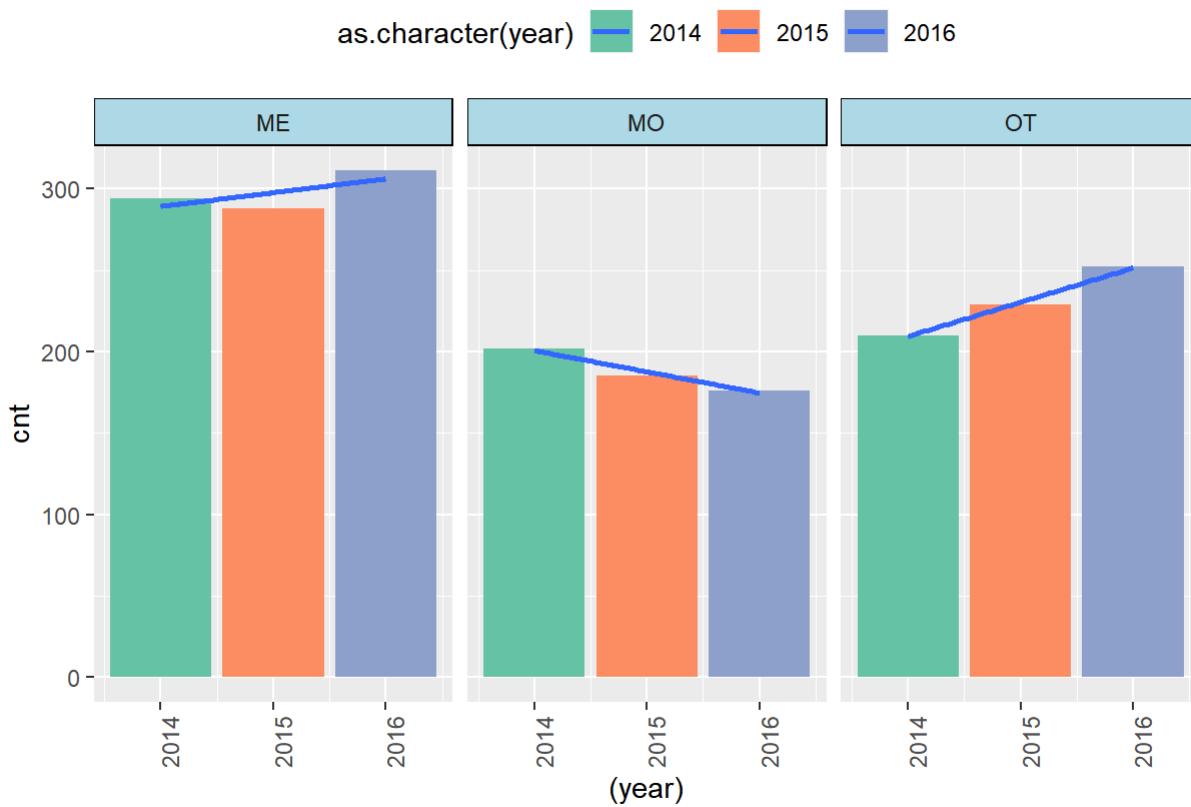
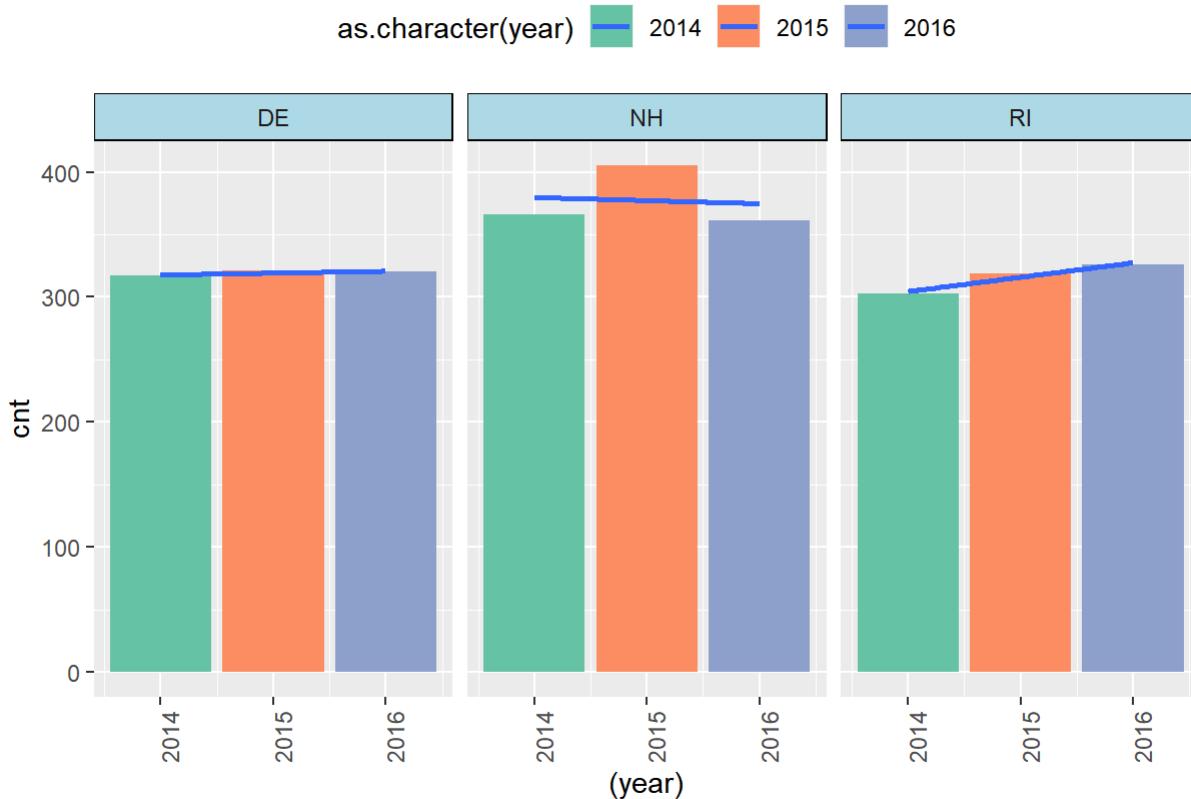
```

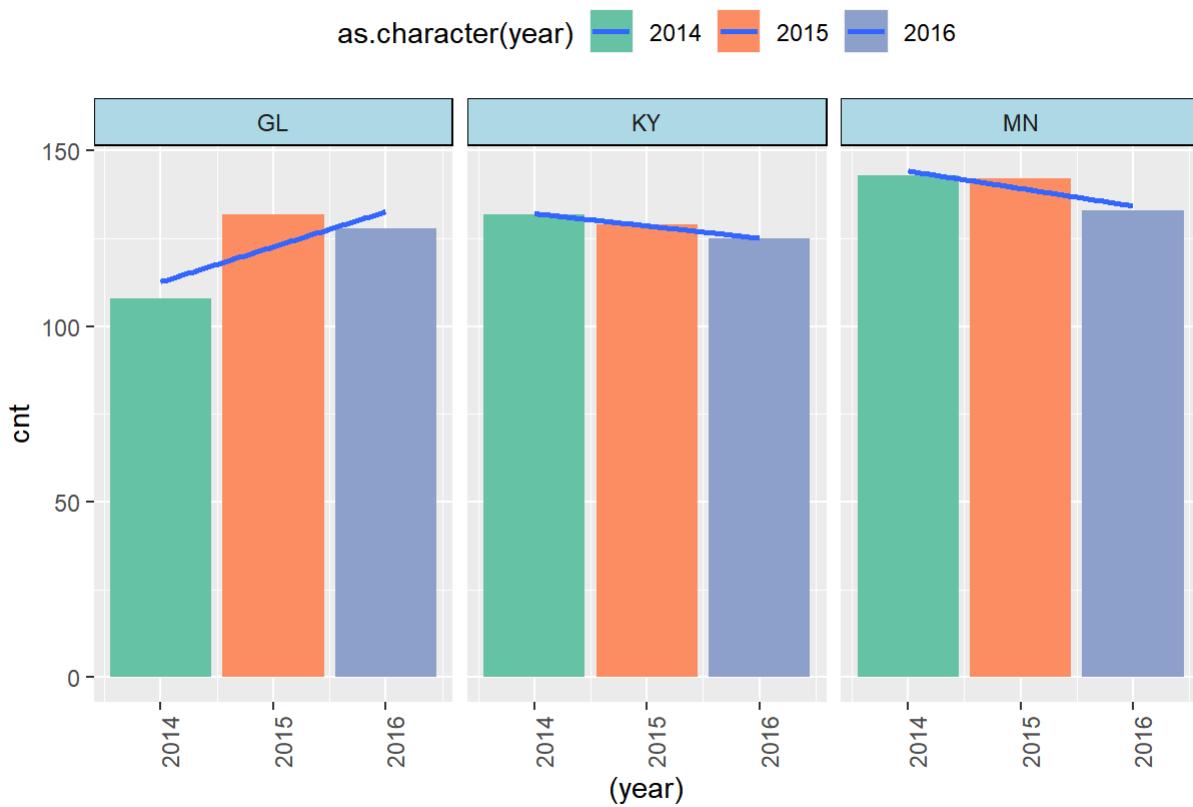
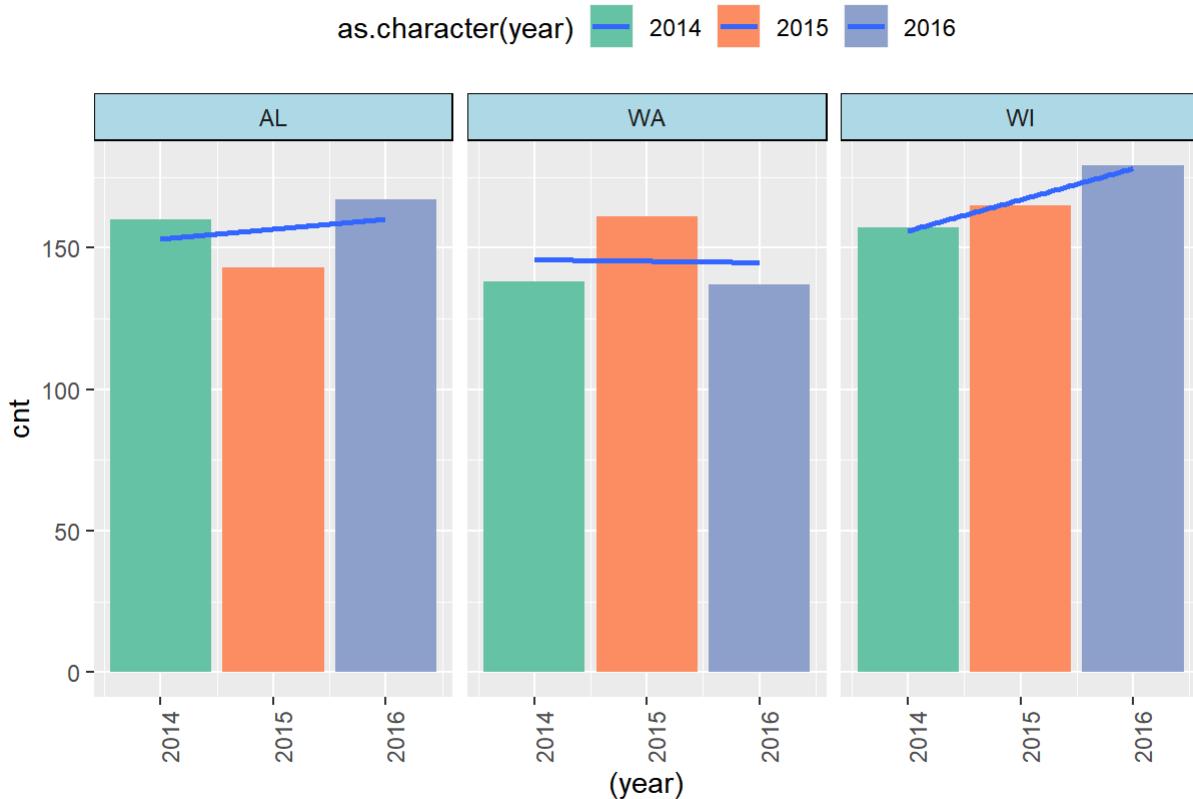











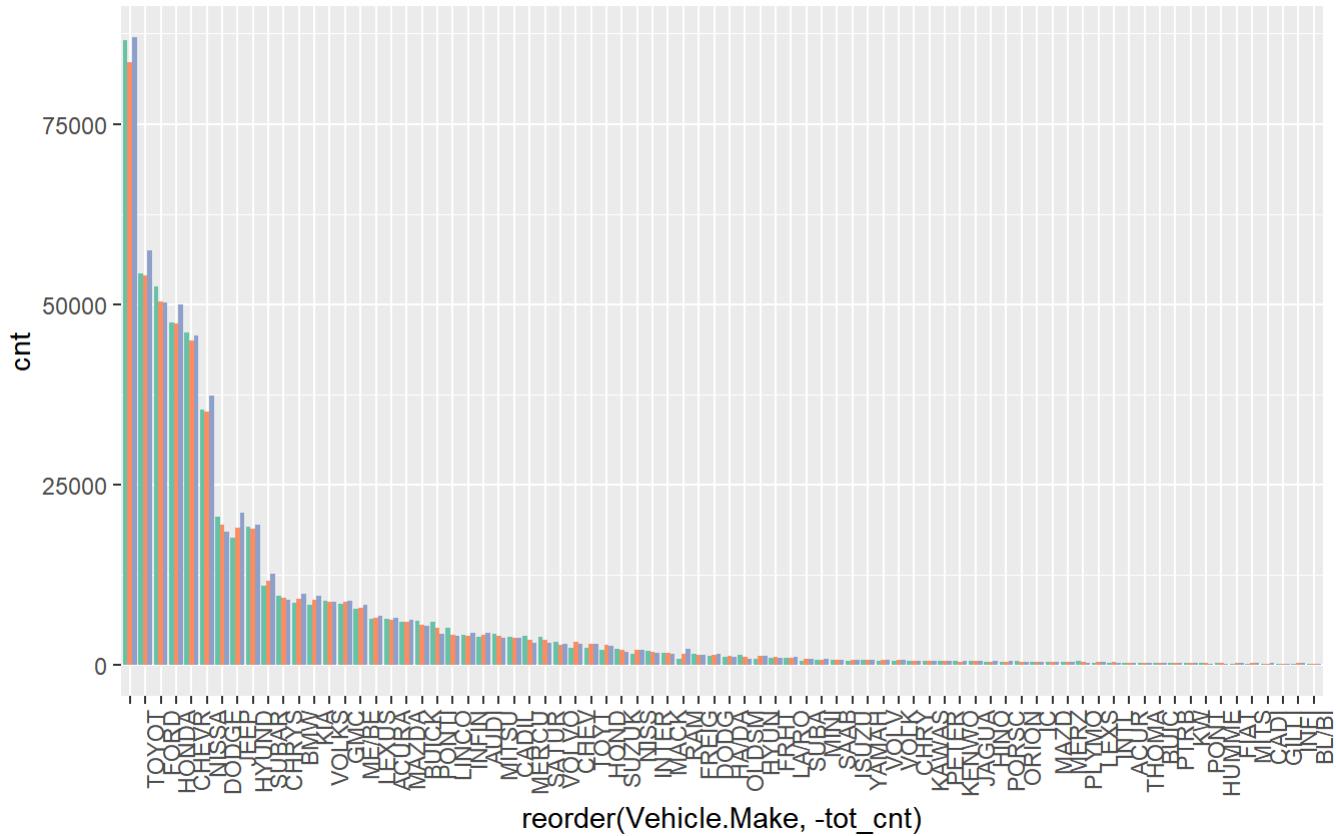




Plot Vehicle.Make YOY breakdown

```
ggplot(my_make_dist[which(my_make_dist$tot_cnt>500),],aes(x=reorder(Vehicle.Make,-tot_cnt), y=cnt,fill=as.character(year)))+geom_bar(stat="identity", position = position_dodge(preserve = "total"))+ theme(legend.position = "top",axis.text.x = element_text(angle = 90, hjust = 1))+ scale_fill_brewer(palette = "Set2")
```

as.character(year) 2014 2015 2016

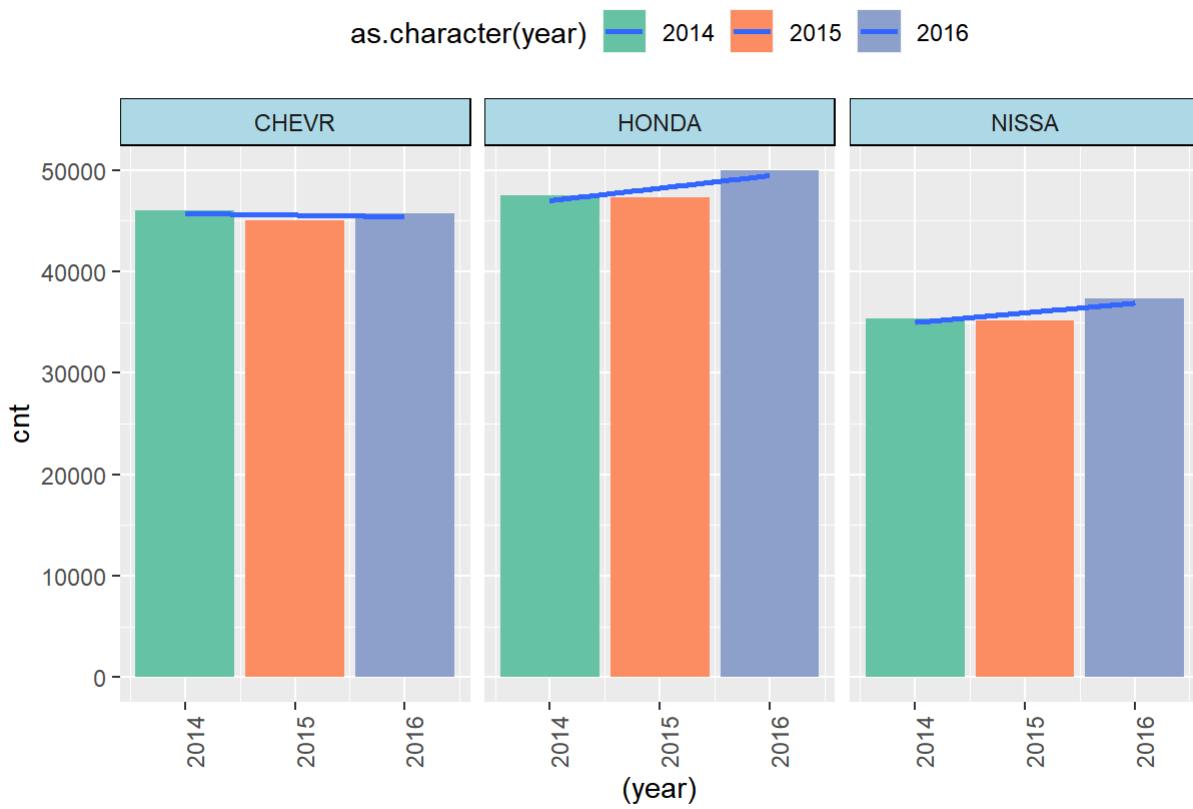
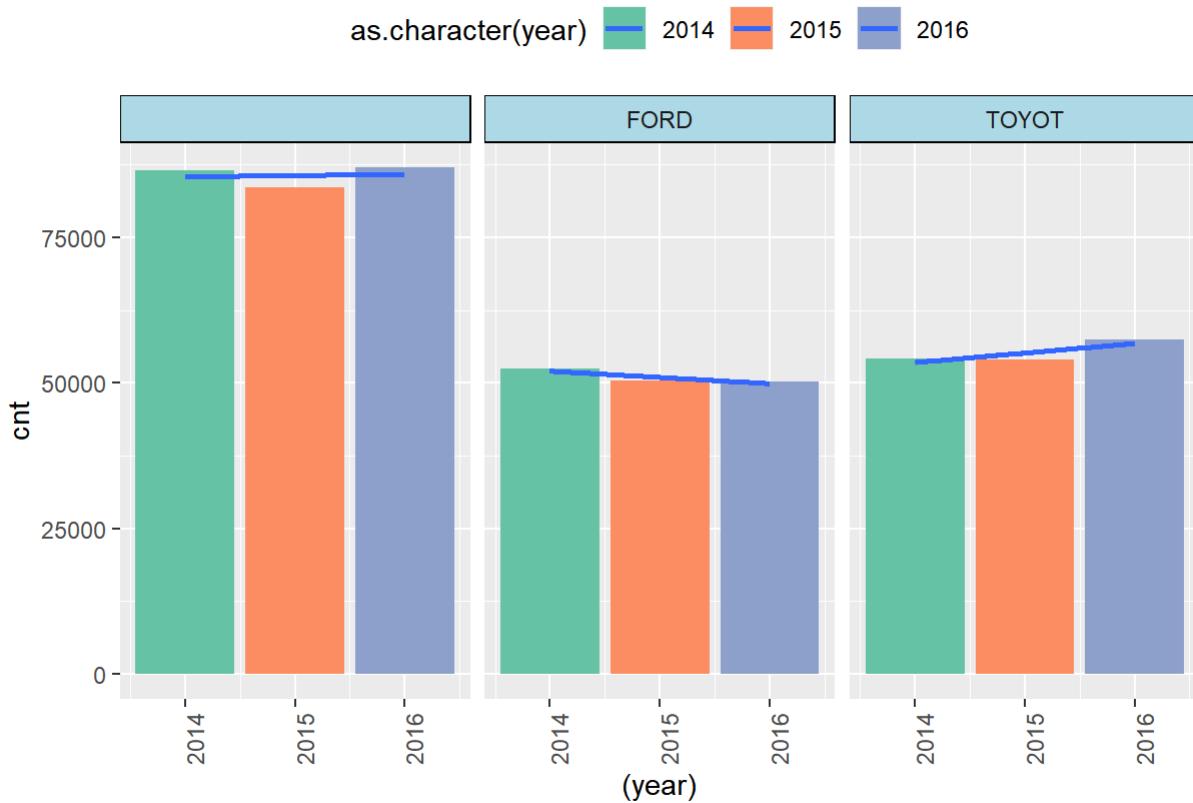


```

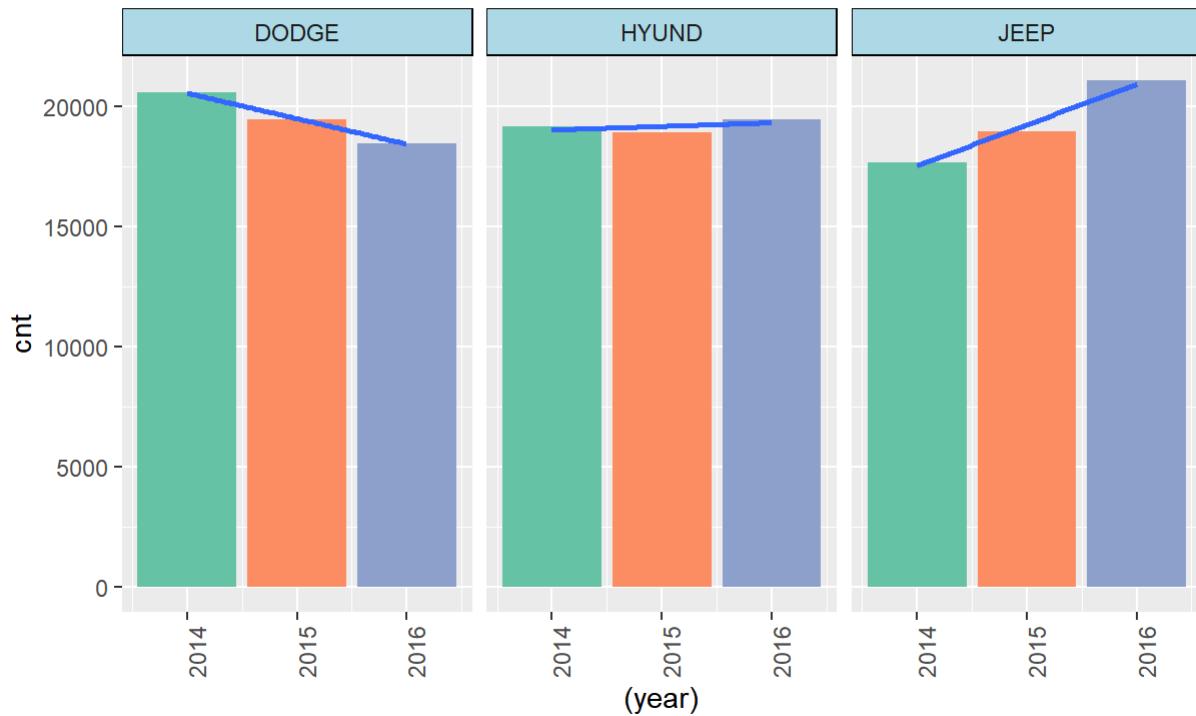
seq <- as.integer(nrow(my_make_dist)[which(my_make_dist$tot_cnt>300 & my_make_dist$yr_cnt==3),])/9
)+1

for (i in 1:seq)
{
  frm <-((i-1)*9)+1
  to<-((i)*9)
print(ggplot(my_make_dist[frm:to,],aes(x=(year), y=cnt,fill=as.character(year)), na.rm=TRUE) +geom_bar(stat="identity") + theme(legend.position = "top",axis.text.x = element_text(angle = 90,
  hjust = 1),strip.background = element_rect(fill="lightblue", colour="black",size=0.5),plot.margin = unit(c(1,1,1,1),"cm"))+ scale_fill_brewer(palette = "Set2") + facet_wrap(~Vehicle.Make, ncol=5, scales = "free_x") + geom_smooth(method = lm, se = FALSE,aes(group = 1)))
}

```

as.character(year) 2014 2015 2016



as.character(year) 2014 2015 2016

