

Japan Import Export

```
library(ggplot2)
## Warning: package 'ggplot2' was built under R version 3.2.4

library(readr)
## Warning: package 'readr' was built under R version 3.2.5

library(plyr)
library(plotly)
## Warning: package 'plotly' was built under R version 3.2.5

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##   last_plot

## The following object is masked from 'package:graphics':
##
##   layout

library(dplyr)
## Warning: package 'dplyr' was built under R version 3.2.5

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

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

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

library(tidyr)
## Warning: package 'tidyr' was built under R version 3.2.5

country_df<-read.csv(file="E:/Capstone Project/japan
dataset/Dataset_new/country_eng.csv", header = TRUE, sep = ",",
stringsAsFactors = F, colClasses = c("Country"="factor"))

hs2_df<-read.csv(file = "E:/Capstone Project/japan
```

```

dataset/Dataset_new/hs2_eng.csv", header = TRUE, sep = "," , stringsAsFactors
= F, colClasses = c("hs2"="factor"))

hs4_df<-read.csv(file = "E:/Capstone Project/japan
dataset/Dataset_new/hs4_eng.csv", header = TRUE,sep = ",", stringsAsFactors =
F, colClasses = c("hs4"="factor"))

hs6_df<-read.csv(file = "E:/Capstone Project/japan
dataset/Dataset_new/hs6_eng.csv", header = TRUE,sep = ",", stringsAsFactors =
F, colClasses = c("hs6"="factor"))

hs9_df<-read.csv(file = "E:/Capstone Project/japan
dataset/Dataset_new/hs9_eng.csv", header = TRUE,sep = ",", stringsAsFactors =
F, colClasses = c("hs9"="factor"))

year_latest_df<-read.csv(file = "E:/Capstone Project/japan
dataset/Dataset_new/year_latest.csv", header = TRUE,sep = ",",
stringsAsFactors = F, colClasses = c("Country"="factor", "hs2"="factor",
"hs4"="factor", "hs6"="factor", "hs9"="factor"))

ym_latest_df<-read.csv(file = "E:/Capstone Project/japan
dataset/Dataset_new/ym_latest.csv", header = TRUE,sep = ",", stringsAsFactors
= F, colClasses = c("Country"="factor", "hs2"="factor", "hs4"="factor",
"hs6"="factor", "hs9"="factor"))

# Check if there is any null value in datasets
which(is.na(country_df))

## integer(0)

which(is.na(hs2_df))

## integer(0)

which(is.na(hs4_df))

## integer(0)

which(is.na(hs6_df))

## integer(0)

which(is.na(year_latest_df))

## integer(0)

which(is.na(ym_latest_df))

## integer(0)

# there is no null values in datasets, so futher process merging the
datasets.

## Merge the country and area codes
merge_year_latest<-join(country_df,year_latest_df,by="Country")
merge_year_latest<-join(merge_year_latest, hs2_df, by= "hs2")

```

```

merge_year_latest<-join(merge_year_latest, hs4_df, by= "hs4")
merge_year_latest<-join(merge_year_latest, hs6_df, by= "hs6")
merge_year_latest<-join(merge_year_latest, hs9_df, by="hs9")
year_latest_df<-merge_year_latest
rm(merge_year_latest)

merge_ym_latest<-join(country_df,ym_latest_df,by="Country")
merge_ym_latest<-join(merge_ym_latest, hs2_df, by= "hs2")
merge_ym_latest<-join(merge_ym_latest, hs4_df, by= "hs4")
merge_ym_latest<-join(merge_ym_latest, hs6_df, by= "hs6")
merge_ym_latest<-join(merge_ym_latest, hs9_df, by="hs9")
ym_latest_df<-merge_ym_latest
rm(merge_ym_latest)

## Variables transformation
## Create new month and derive exp_imp flg
ym_latest_df$month_char<-month.abb[ym_latest_df$month]
ym_latest_df$month_char<-factor(ym_latest_df$month_char, levels = month.abb)

addExportFlg<- function(datasets){
  datasets$type<-ifelse(datasets$exp_imp == 1, "Export", "Import")
  return(datasets)
}

smp1<-function(dataset){
  return(head(dataset,2))
}

year_latest_df<-addExportFlg(year_latest_df)
ym_latest_df<-addExportFlg(ym_latest_df)

## Drop some columns
drops<-c("exp_imp", "Country")
year_latest_df<-year_latest_df[ , ! (names(year_latest_df)%in% drops)]

drops<-c("exp_imp", "Country", "month")
ym_latest_df<-ym_latest_df[ , ! (names(ym_latest_df)%in% drops)]

rm(drops)

## Sample after Transformation datasets
head(year_latest_df,2)

##      Country_name Area Year Unit1 Unit2 QY1  QY2  VY hs2  hs4  hs6
## 1 Republic_of_Korea Asia 1988      KG  0 25659 38239 01 0106 010600
## 2 Republic_of_Korea Asia 1988      KG  0   49  265 02 0208 020890
##      hs9      hs2_name
## 1 010600900      Live animals.
## 2 020890900 Meat and edible meat offal.
##                                     hs4_name hs6_name
## 1                                     Other_live_animals <NA>
## 2 Other_meat_and_edible_meat_offal__fresh__chilled_or_frozen Other
##      hs9_name type

```

```
## 1      <NA> Export
## 2      <NA> Export

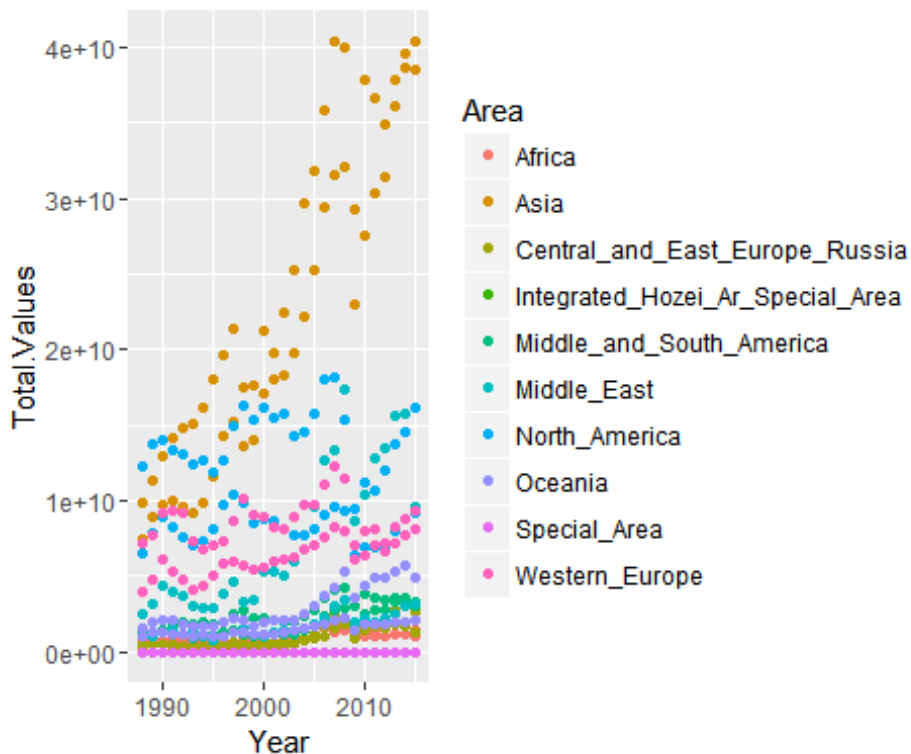
head(ym_latest_df,2)

##      Country_name Area Year Unit1 Unit2 Q1      Q2      Value hs2  hs4
## 1 Republic_of_Korea Asia 2015      KG  0  9088903 21026081  00 0000
## 2 Republic_of_Korea Asia 2015      KG  0 134125685 24764709  00 0000
##      hs6      hs9      hs2_name      hs4_name      hs6_name
## 1 000000 000000190 re_import_export re_import_export re_import_export
## 2 000000 000000190 re_import_export re_import_export re_import_export
##      hs9_name month_char  type
## 1 re_export_ohters      May Export
## 2 re_export_ohters      Feb Export
```

Trade as Year wise

```
exp_summary_year_Export<-
aggregate(as.numeric(year_latest_df$VY),by=list(year_latest_df$Year
,year_latest_df$type, year_latest_df$Area),FUN = sum)
colnames(exp_summary_year_Export)<-c("Year","Type", "Area","Total.Values")

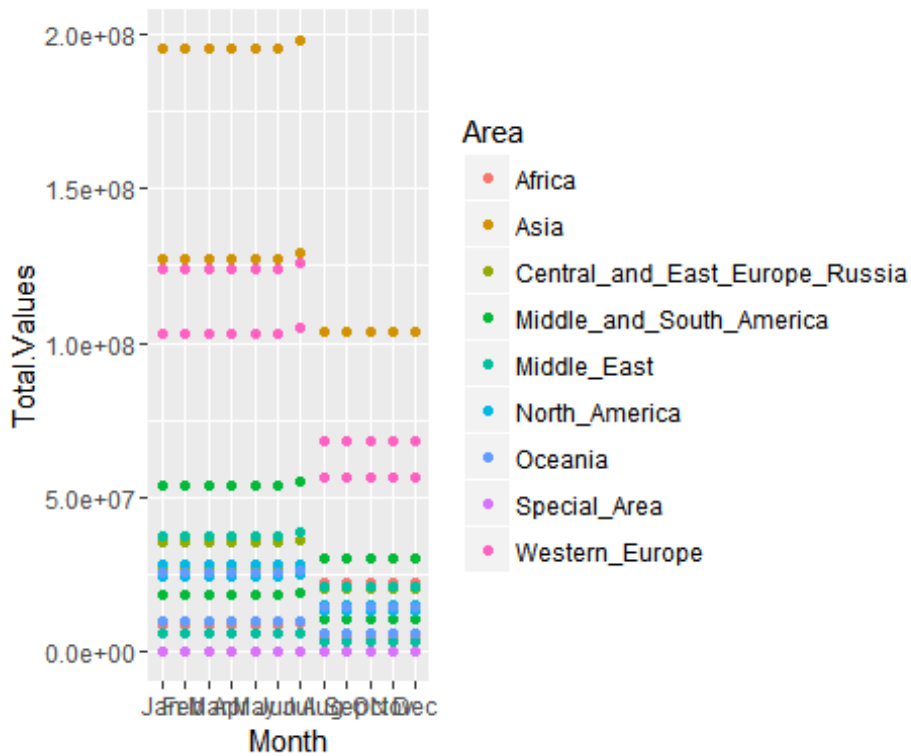
g1<-ggplot(exp_summary_year_Export, aes(x=Year, y= Total.Values ,color=
Area))
g1 + geom_point()
```



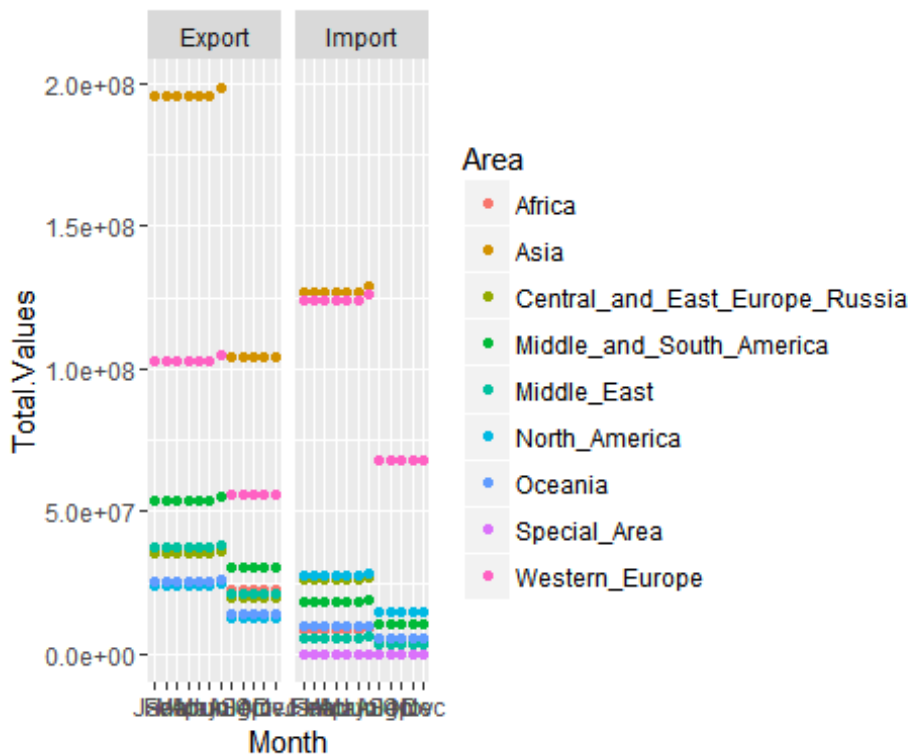
Trade as Month wise

```
exp_summary_ym_Import<-
aggregate(as.numeric(ym_latest_df$Year),by=list(ym_latest_df$month
,ym_latest_df$type, ym_latest_df$Area),FUN = sum)
colnames(exp_summary_ym_Import)<-c("Month","Type", "Area","Total.Values")
```

```
g1<-ggplot(exp_summary_ym_Import, aes(x=Month, y= Total.Values ,color= Area))
g1 + geom_point()
```



```
g1 + geom_point(aes(color = Area)) + facet_wrap(~Type, ncol = 2 )
```



```
# Trade with Middle_East Area
middle_east_df<-filter(year_latest_df, Area == "Middle_East")
which(is.na(middle_east_df$hs2_name))
```

```
## integer(0)

#grouped_by_hs2<-group_by(middle_east_df, Country_name, hs2, Year, VY)
#summarize(grouped_by_hs2, mean=mean(VY))

# Summarize data
#vys<-
aggregate(as.numeric(grouped_by_hs2$VY), by=list(grouped_by_hs2$Year, grouped_by_hs2$Country_name, grouped_by_hs2$hs2), FUN=sum)
#colnames(vys)<-c("Year", "Country_name", "hs2", "VY")

# join vys with hs2_data
#vys<-join(vys, hs2_df, by = "hs2")
#ggplot(vys, aes(x=Year, y=VY)) + geom_line() + facet_wrap(~hs2_name)

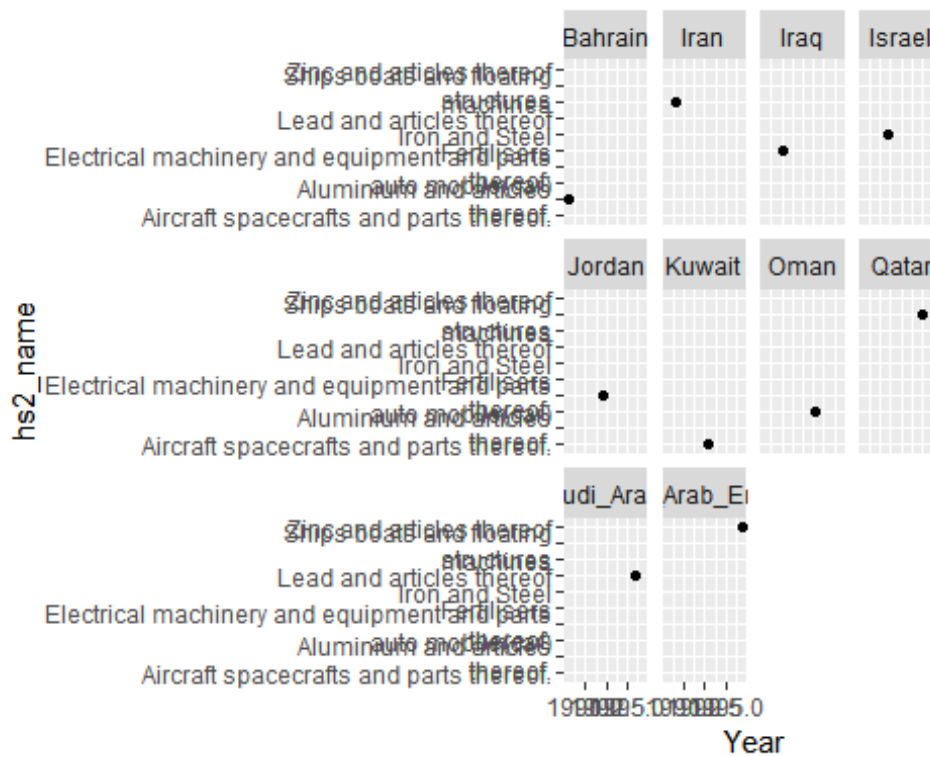
#main_goods<-filter(vys, VY >= 0.4 )

#hs2_names<-unique(main_goods, main_goods$hs2_name, incomparables = FALSE)

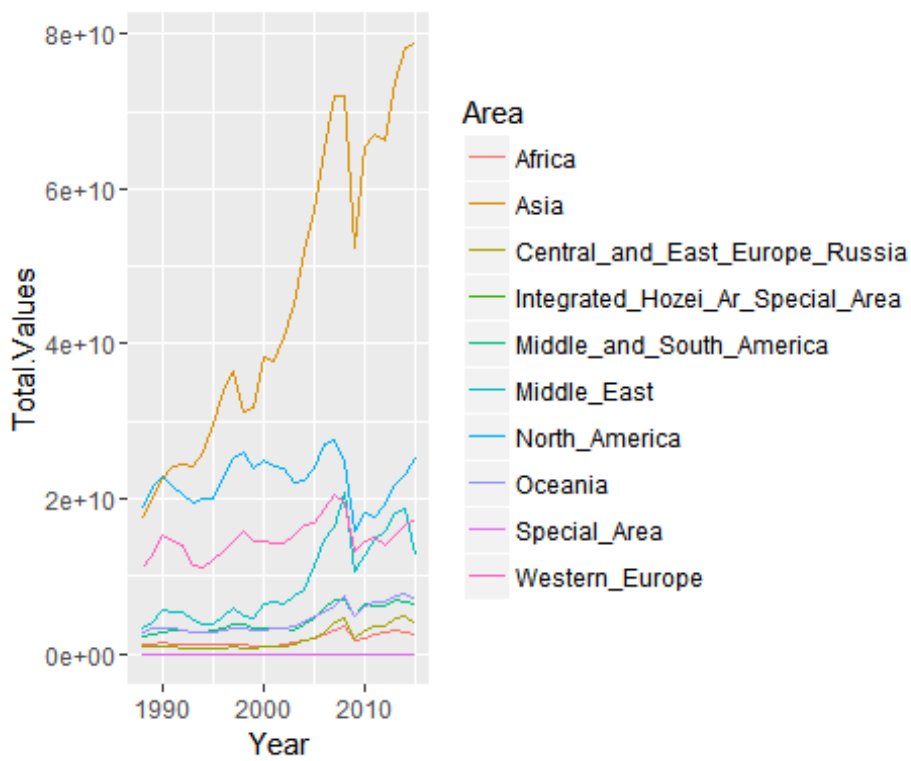
#for(hs2_name in hs2_names){
#  #boxplot(VY~Year, data=hs2_names, xlab="Year", ylab="VY", col=topo.colors(4))
#}
#ggplot(hs2_names, aes(x=Year, y=VY, color=Country_name)) + geom_line() +
facet_wrap(~hs2_name)

df<-
data.frame(Country_name=c("Bahrain", "Iran", "Iraq", "Israel", "Jordan", "Kuwait",
"Oman", "Qatar",
"Saudi_Arabia", "United_Arab_Emirates"), hs2_name = c("Aluminium and articles
thereof.", "machines", "Fertilisers", "Iron and Steel", "Electrical machinery and
equipment and parts
thereof;", "Aircraft spacecrafts and parts thereof.", "auto mobile(car)", "Ships
boats and floating
structures", "Lead and articles thereof", "Zinc and articles thereof"), Year =
c(1988:1997))

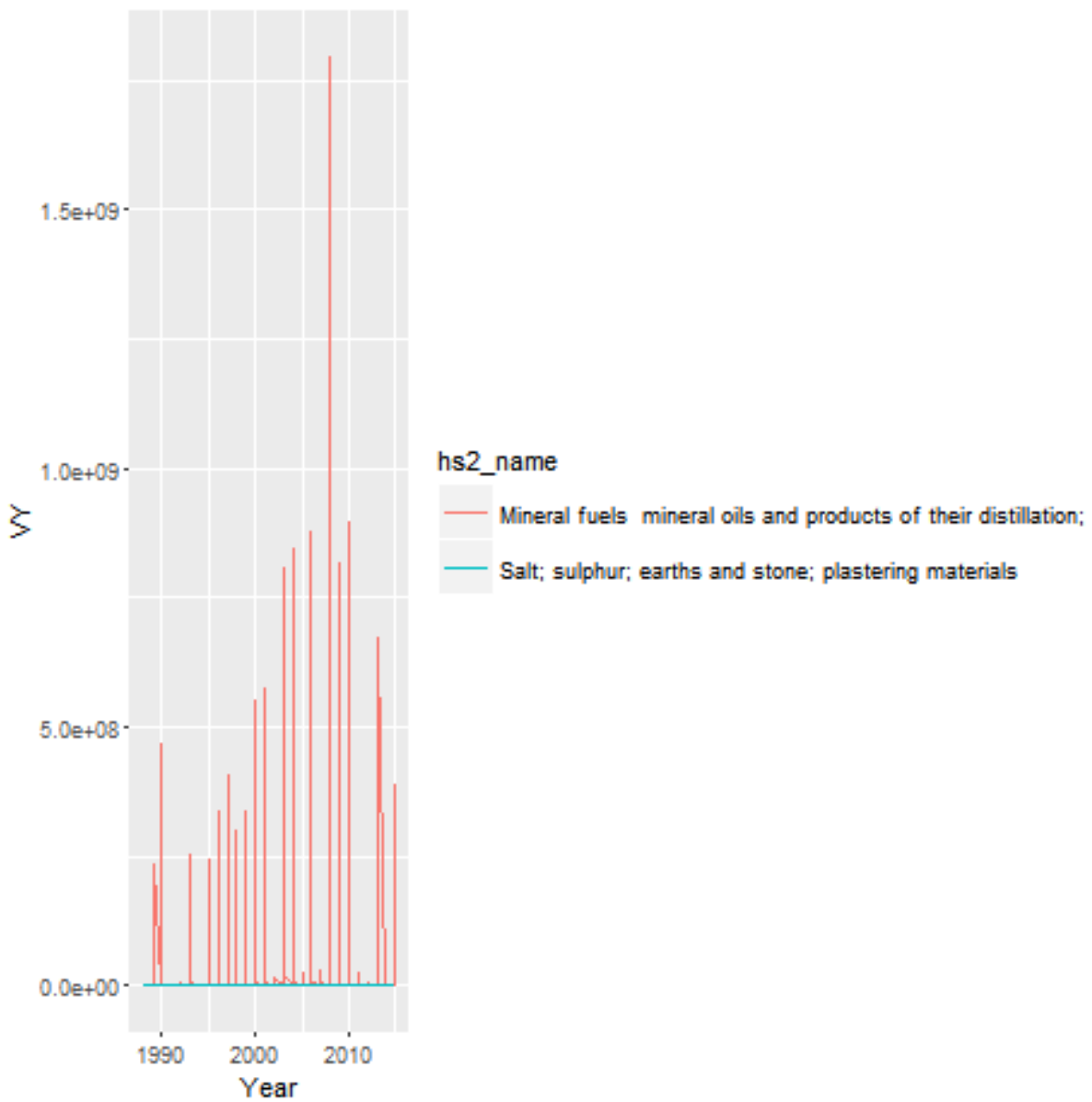
main_goods<- unique(df[c("Country_name", "hs2_name", "Year")])
#merge_vys<-join(vys, main_goods, by="Country_name")
ggplot(main_goods, aes(x=Year, y=hs2_name)) + geom_point() +
facet_wrap(~Country_name)
```



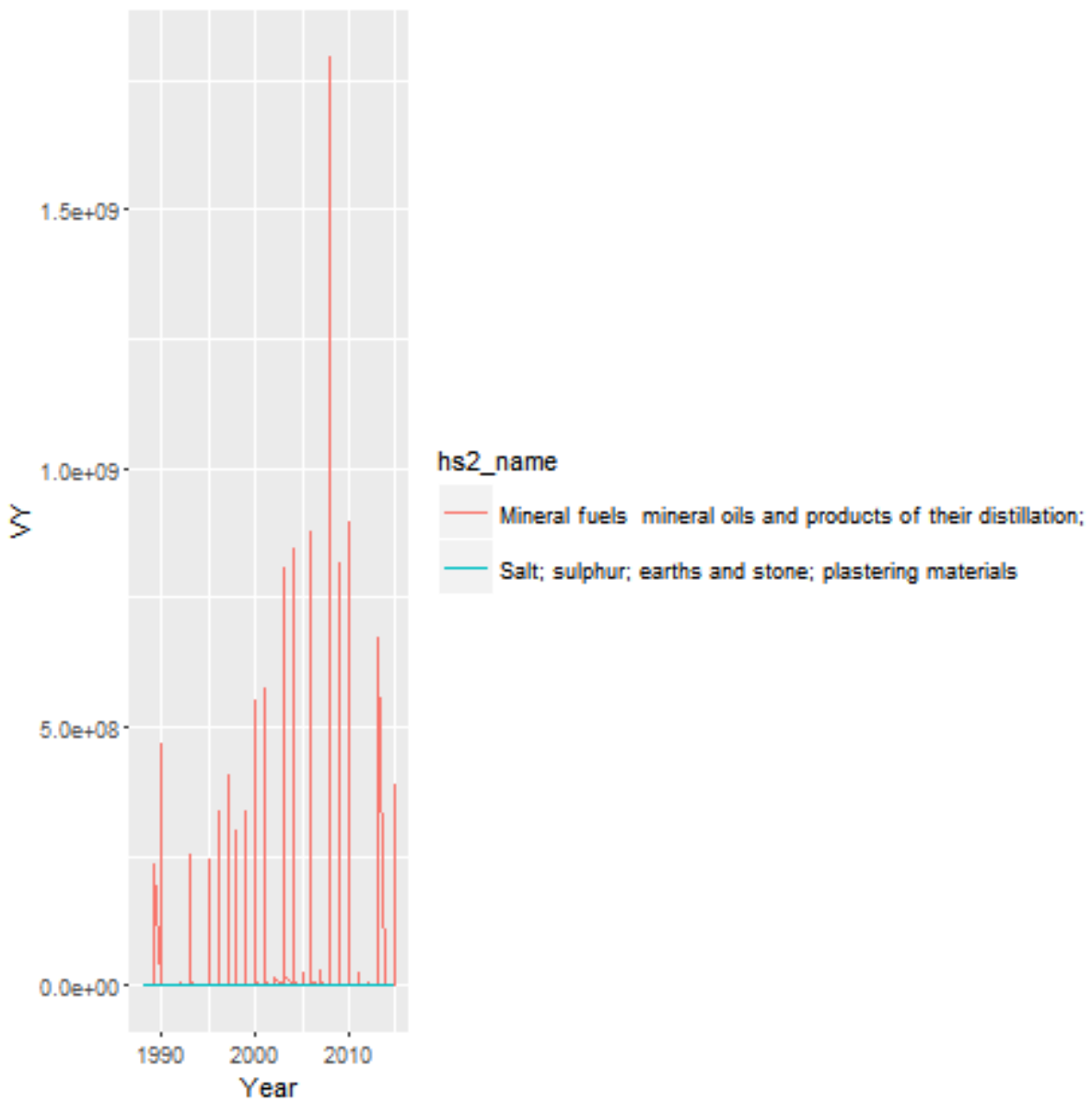
```
# Trade with Areas
areas<-unique(country_df$Area)
grouped_by_area<-group_by(year_latest_df,Year,Area)
vys<-aggregate(as.numeric(grouped_by_area$VY),by =
list(grouped_by_area$Year,grouped_by_area$Area),FUN = sum)
colnames(vys)<-c("Year","Area","Total.Values")
#boxplot(Area~Total.Values,data=vys,xlab="Area",ylab="Total Values")
ggplot(vys,aes(x=Year,y=Total.Values,color=Area)) +
geom_line()#geom_freqpoly(aes(group = Area, color = Area))
```



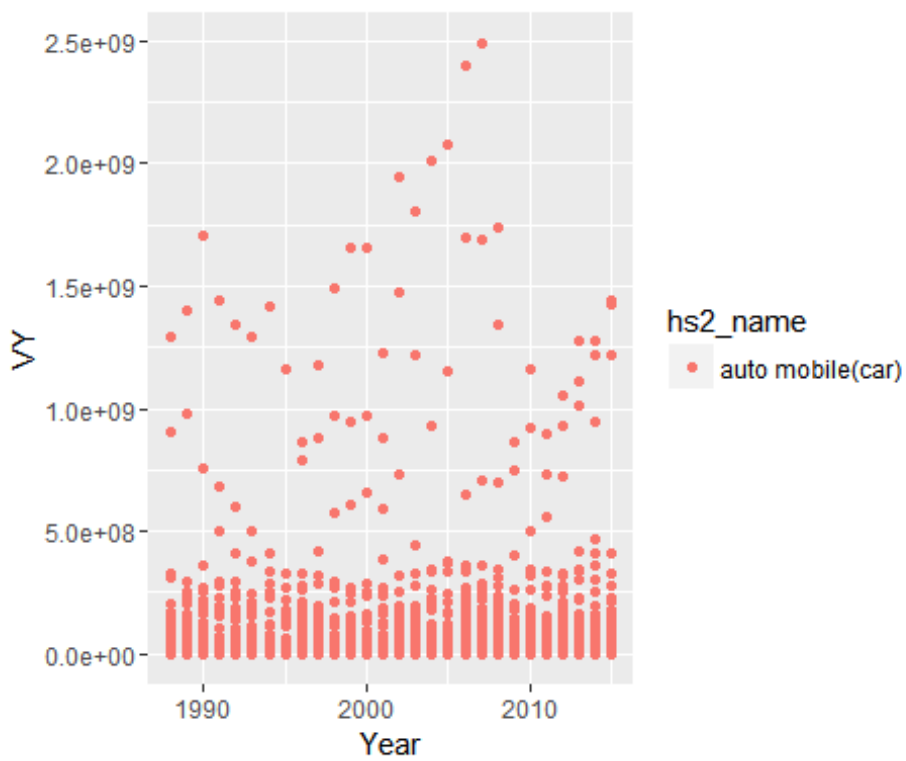
```
# Trade with Iran(Middle East Area) for particular product
df<-filter(middle_east_df, Country_name == "Iran", hs2==c("25", "27"))
grouped_by_df<-group_by(df, Year, VY, hs2_name)
ggplot(grouped_by_df, aes(x=Year, y=VY, color=hs2_name)) +
geom_line(aes(color=hs2_name))
```

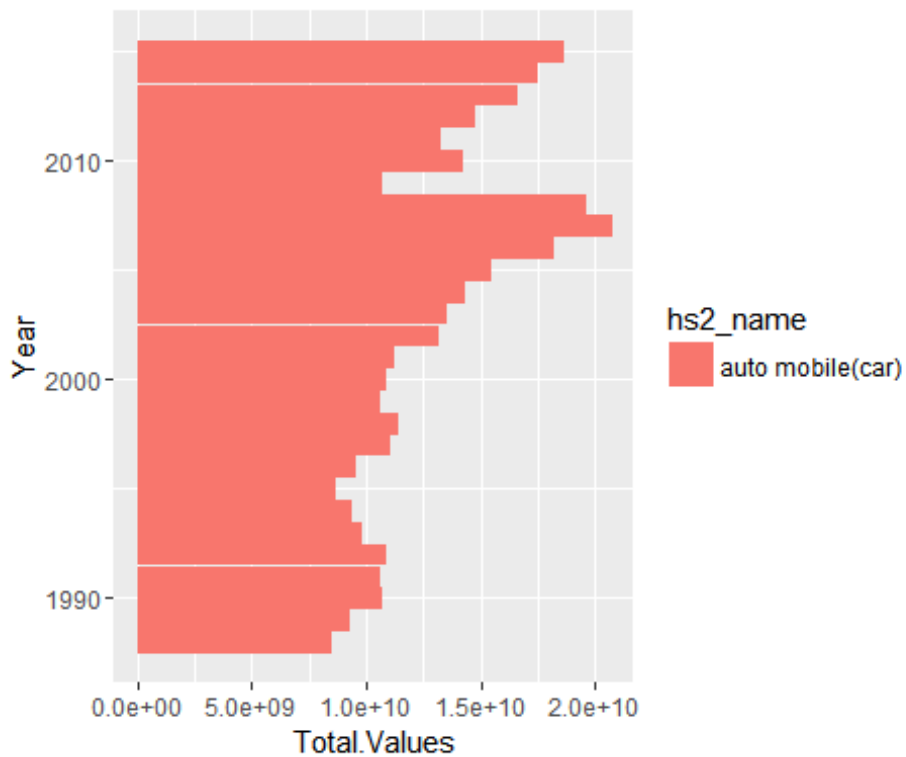
```
vys<-
aggregate(as.numeric(grouped_by_df$VY),by=list(grouped_by_df$Year,grouped_by_
df$hs2_name),FUN=sum)
colnames(vys)<-c("Year","hs2_name","Total.Values")
# Visulization View
ggplot(vys,aes(x=Year,y=Total.Values,color=hs2_name)) + geom_line()
```



```
# Trade of automobiles(car) with Japan
iran_df<-filter(year_latest_df,hs2 == "87")
grouped_by_iran<-group_by(iran_df,Year,VY,hs2_name)
ggplot(grouped_by_iran,aes(x=Year,y=VY,color=hs2_name)) + geom_point()
```



```
vys<-
aggregate(as.numeric(grouped_by_iran$VY),by=list(grouped_by_iran$Year,grouped
_by_iran$hs2_name),FUN = sum)
colnames(vys)<-c("Year","hs2_name","Total.Values")
ggplot(vys,aes(x=Year,y=Total.Values,color=hs2_name)) +
geom_bar(stat="Identity",aes(fill= hs2_name)) + coord_flip()
```



```
ggplot(vys,aes(x=Year,y=Total.Values,color = hs2_name)) + geom_line()
```

