Case_2

Alexandre Martins Fontes

04/07/2020

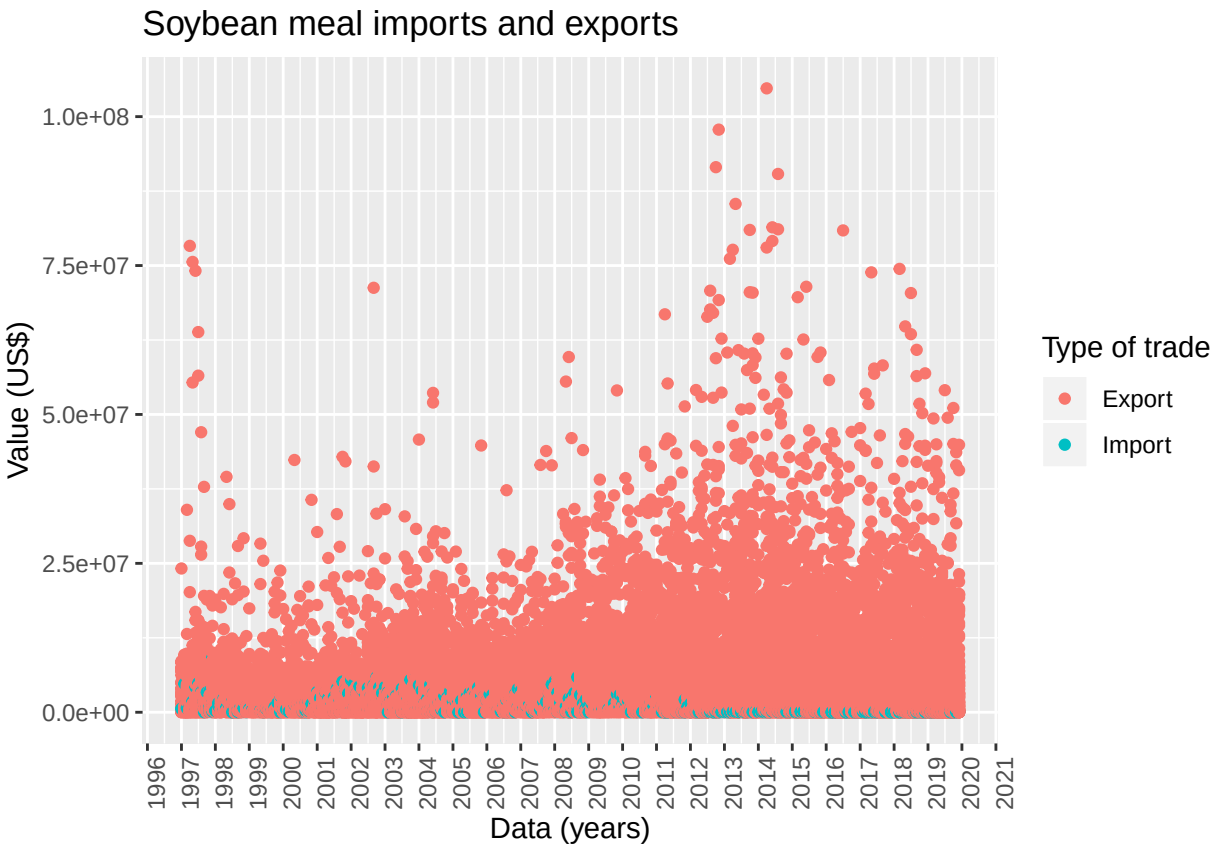
1)Show the evolution of total monthly and total annual exports from Brazil (all states and to everywhere) of 'soybeans', 'soybean oil' and 'soybean meal'

Imports were also included in plots for a supperficial comparation between magnetudes.

Data manipulation for Soybeans by year plot

All the data manipulation are avaliable and also can be explored/reproduced in the file case_2.R

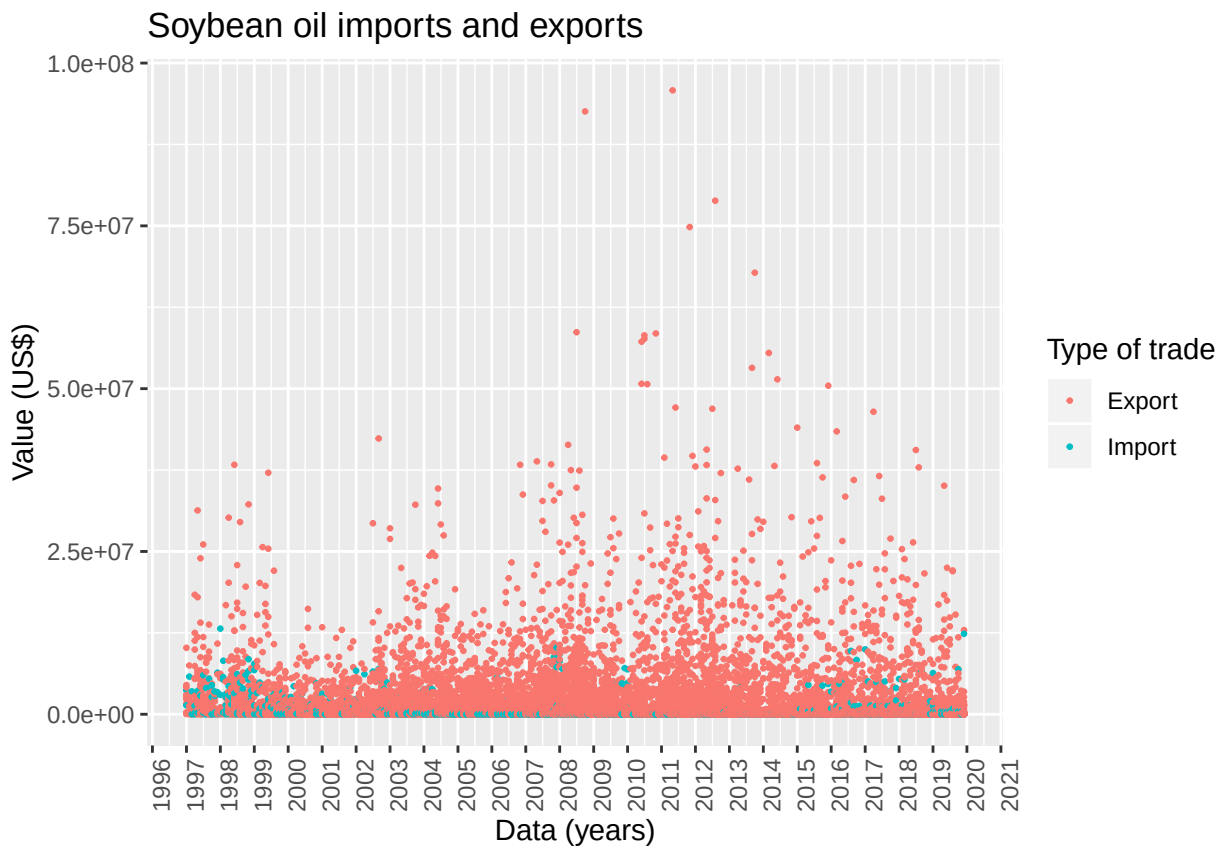
```
knitr::opts_chunk$set(echo = T)

#####CODE WITH DATA MANIPULATION#####
#soybean-meal by year
rm(list = 'data2')

data2 = subset(data, data$product=='soybean_meal' )

data2$date = as.Date(data2$date)

p <- ggplot(data2, aes(x = data2$date, y = data2$usd , color = as.factor(data2$type)) ) +
  geom_point() +
  scale_x_date(date_breaks = "years", date_labels = "%Y")+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Soybean meal imports and exports", x = "Data (years)", y = "Value (US$)", color = "Type

p
```



Data manipulation for Soybean meal by year plot

All the data manipulation are avaliable and also can be explored/reproduced in the file case_2.R

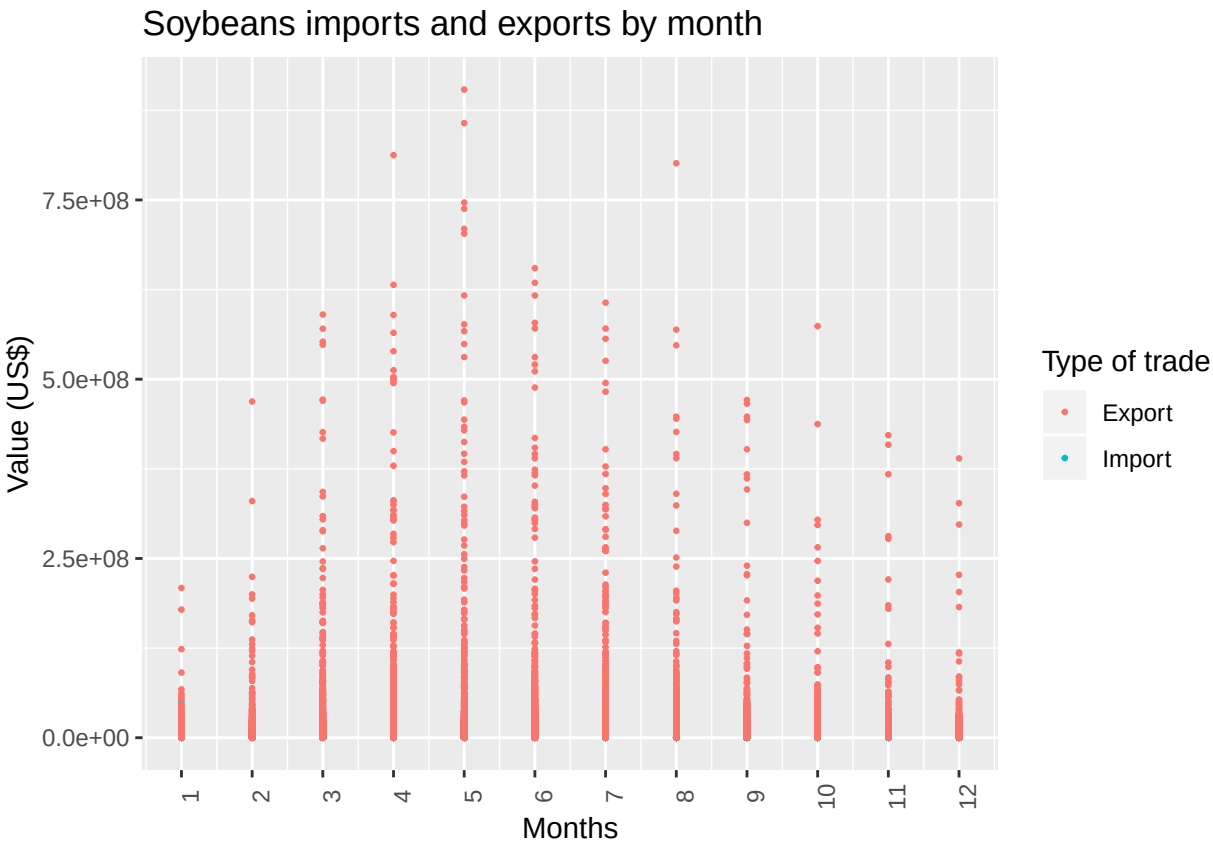
```
knitr::opts_chunk$set(echo = T)

#####CODE WITH DATA MANIPULATION#####
#soybean-oil by year
rm(list = 'data2')

data2 = subset(data, data$product=='soybean_oil' )

data2$date = as.Date(data2$date)

p <- ggplot(data2, aes(x = data2$date, y = data2$usd , color = as.factor(data2$type)) ) +
  geom_point(size = 0.5) +
  scale_x_date(date_breaks = "years", date_labels = "%Y")+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Soybean oil imports and exports", x = "Data (years)", y = "Value (US$)", color = "Type

p
```



Data manipulation for Soybean oil by year plot

All the data manipulation are avaliable and also can be explored/reproduced in the file case_2.R

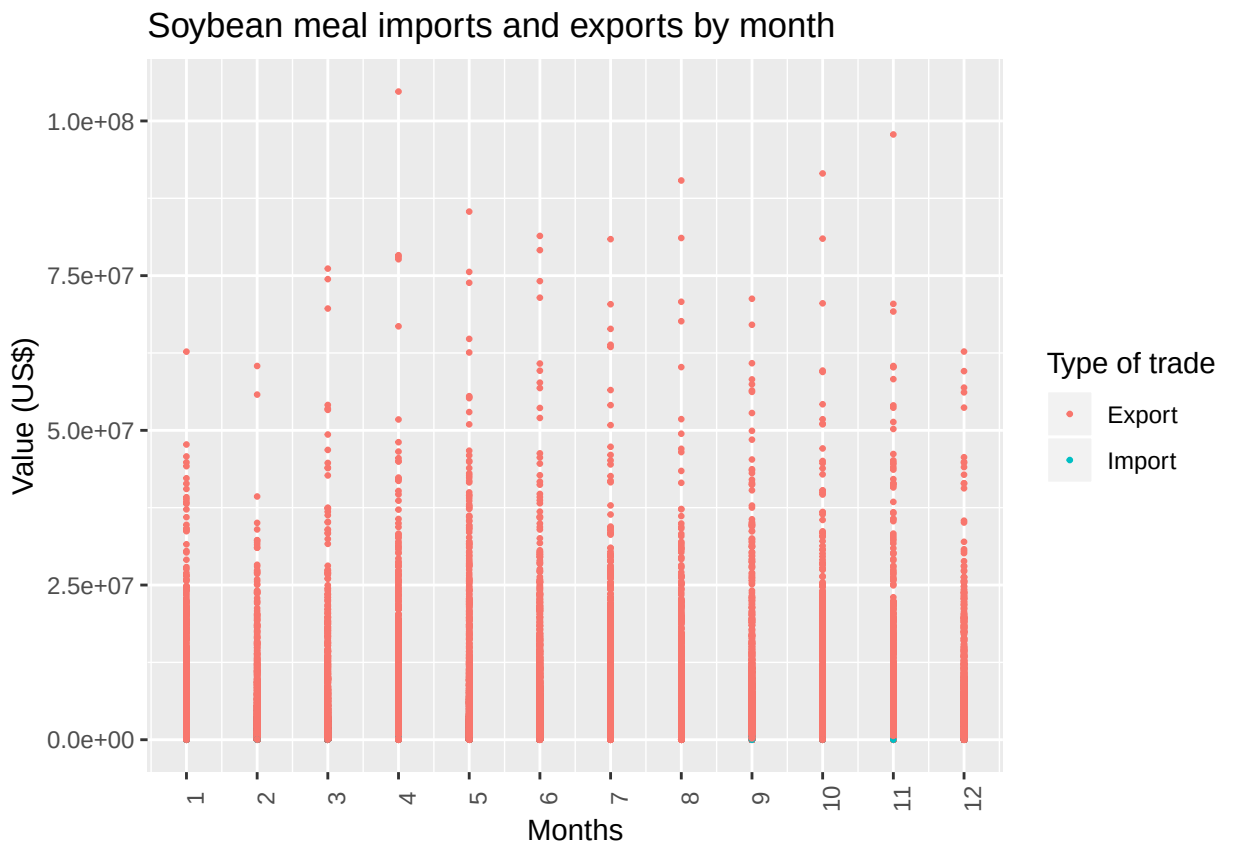
```
knitr::opts_chunk$set(echo = T)

#####CODE WITH DATA MANIPULATION#####
#soybeans by month
rm(list = 'data2')

data2 = subset(data, data$product=='soybeans' )

data2$date = as.numeric(substr ( data2$date, 6,7 ))




p <- ggplot(data2, aes(x = data2$date, y = data2$usd , color = as.factor(data2$type)) ) +
  geom_point(size = 0.5) +
  scale_x_continuous(breaks = seq(1,12, by = 1))+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Soybeans imports and exports by month", x = "Months", y = "Value (US$)", color = "Type

p
```



Data manipulation for soybeans by month plot

All the data manipulation are avaliable and also can be explored/reproduced in the file case_2.R

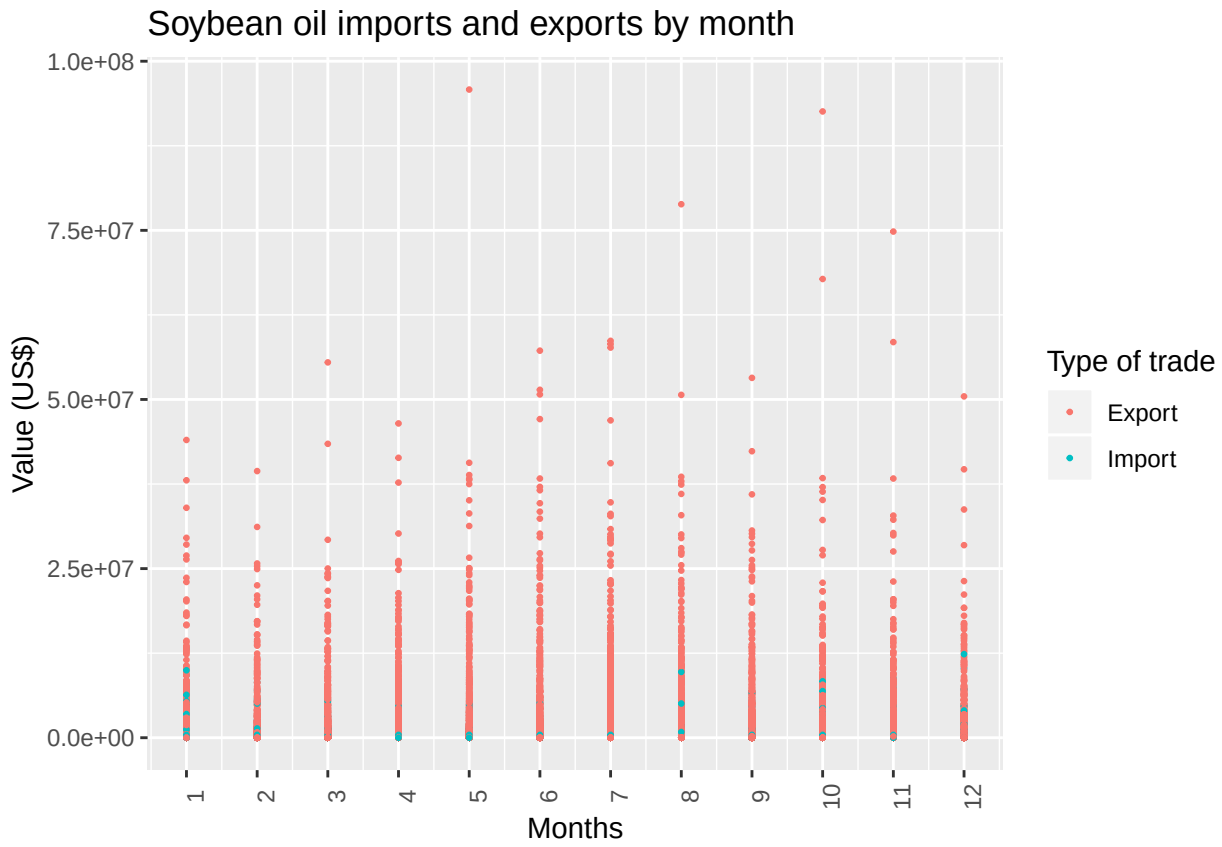
```
knitr::opts_chunk$set(echo = T)

#####CODE WITH DATA MANIPULATION#####
#soybean meal by month
rm(list = 'data2')

data2 = subset(data, data$product=='soybean_meal' )

data2$date = as.numeric(substr ( data2$date, 6,7 ))




p <- ggplot(data2, aes(x = data2$date, y = data2$usd , color = as.factor(data2$type)) ) +
  geom_point(size = 0.5) +
  scale_x_continuous(breaks = seq(1,12, by = 1))+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Soybean meal imports and exports by month", x = "Months", y = "Value (US$)", color = "Ty

p
```



Data manipulation for soybean meal by month plot

All the data manipulation are avaliable and also can be explored/reproduced in the file case_2.R

```
knitr::opts_chunk$set(echo = T)

#####CODE WITH DATA MANIPULATION#####
#soybean oil by month
rm(list = 'data2')

data2 = subset(data, data$product=='soybean_oil' )

data2$date = as.numeric(substr ( data2$date, 6,7 ))




p <- ggplot(data2, aes(x = data2$date, y = data2$usd , color = as.factor(data2$type)) ) +
  geom_point(size = 0.5) +
  scale_x_continuous(breaks = seq(1,12, by = 1))+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Soybean oil imports and exports by month", x = "Months", y = "Value (US$)", color = "Ty

p
```



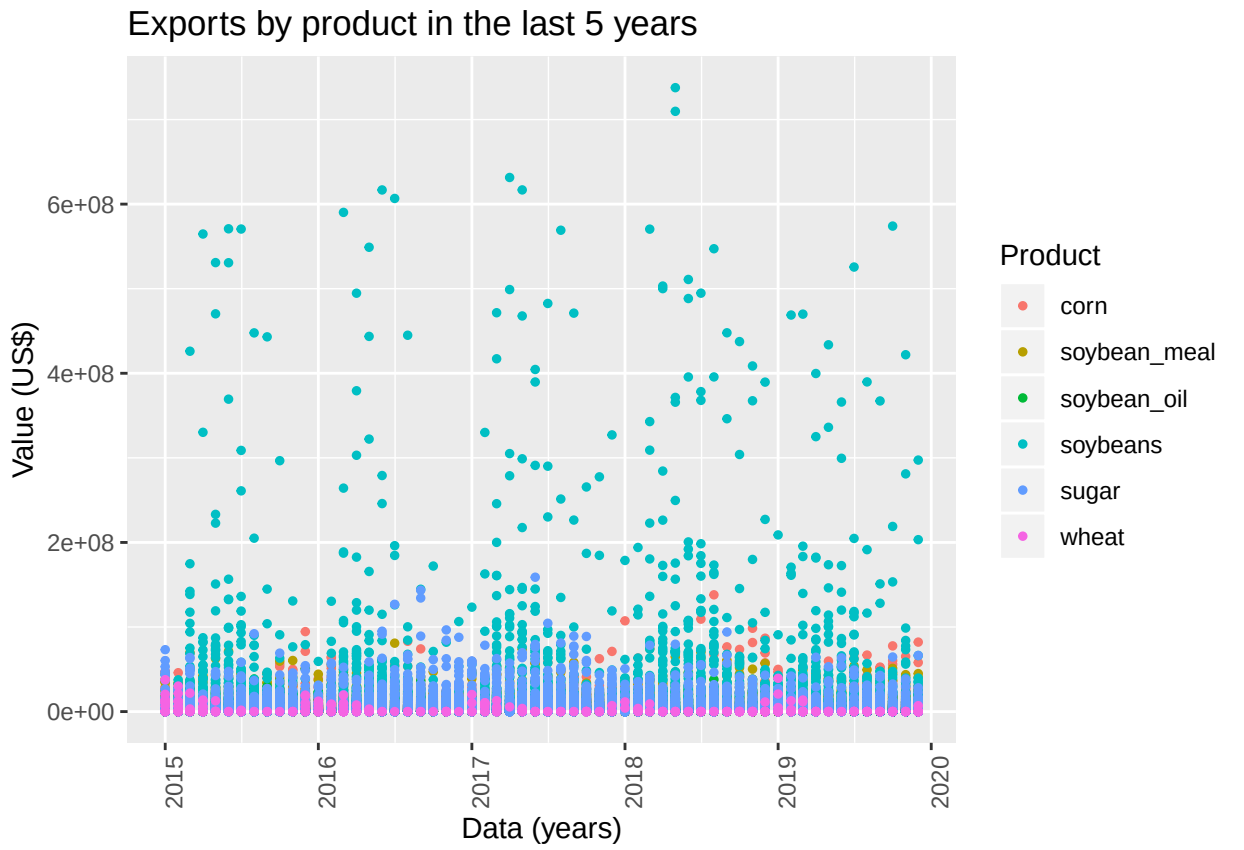
Data manipulation for soybean oil by month plot

All the data manipulation are avaliable and also can be explored/reproduced in the file case_2.R

2) What are the 3 most important products exported by Brazil in the last 5 years?

```
knitr::opts_chunk$set(echo = T)
#3 most important products
rm(list = 'data2')

data2 = subset(data, data$type=='Export' )
data2$date = as.Date(data2$date)
data2= data2[ data2$date > as.Date("2014-12-01"), ]

corn_5 = sum(data2$usd[data2$product=='corn'])
soybean_meal_5 = sum(data2$usd[data2$product=='soybean_meal'])
soybean_oil_5 = sum(data2$usd[data2$product=='soybean_oil'])
soybeans_5 = sum(data2$usd[data2$product=='soybeans'])
sugar_5 = sum(data2$usd[data2$product=='sugar'])
wheat_5 = sum(data2$usd[data2$product=='wheat'])

three_most_important = c(soybeans_5,sugar_5,soybean_meal_5)
p <- ggplot(data2, aes(x = data2$date, y = data2$usd , color = as.factor(data2$product)) ) +
  geom_point(size = 1) +
  scale_x_date(date_breaks = "years", date_labels = "%Y")+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Exports by product in the last 5 years", x = "Data (years)", y = "Value (US$)", color =
```

```
p
```

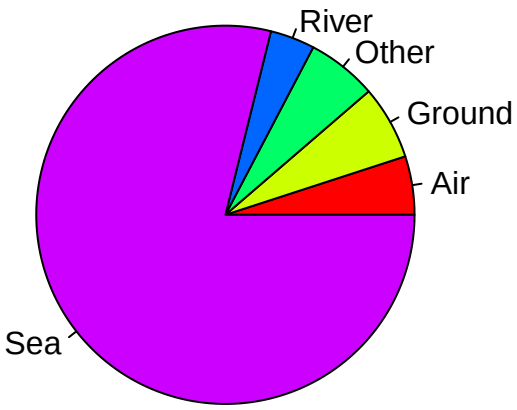


It is clear from the graph that 'soybeans' is the product most responsible for financial movements in the last 5 years.

The second and third most important products are 'sugar' and 'soybean meal'

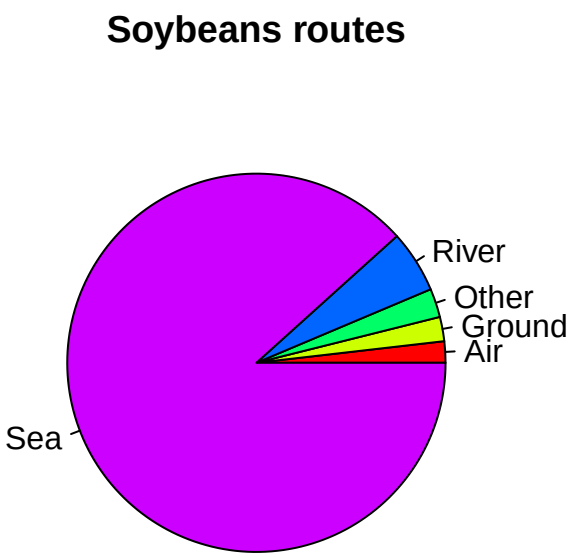
3) What are the main routes through which Brazil have been exporting 'corn' in the last few years? Are there differences in the relative importancem of routes depending on the product?

```
knitr::opts_chunk$set(echo = T)

rm(list = 'data2','soybeans_5','sugar_5','soybean_meal_5','soybean_oil_5','corn_5')

data2 = subset(data, data$type=='Export' )
data2$date = as.Date(data2$date)
data2= data2[ data2$date > as.Date("2014-12-01"), ]
cor = table(data2$route[data2$product=='corn'])
pie(table(data2$route[data2$product=='corn']), main = "Corn routes", col = rainbow(length(cor)))
```

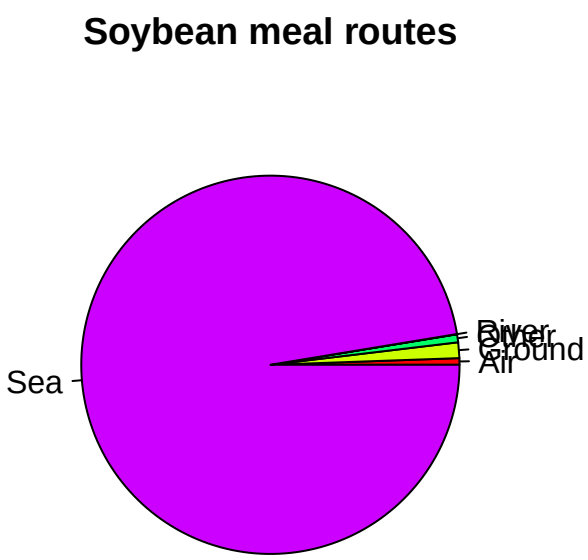


Corn is exported in the last years mostly by sea

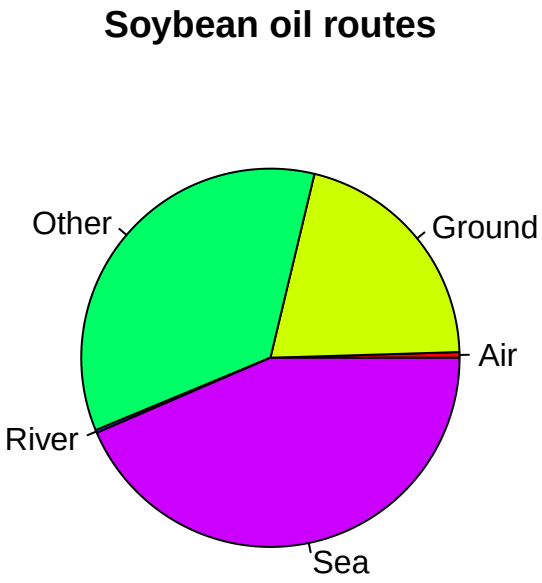
```
knitr::opts_chunk$set(echo = T)

pie(table(data2$route[data2$product=='soybeans']), main = "Soybeans routes", col = rainbow(length(cor))
```



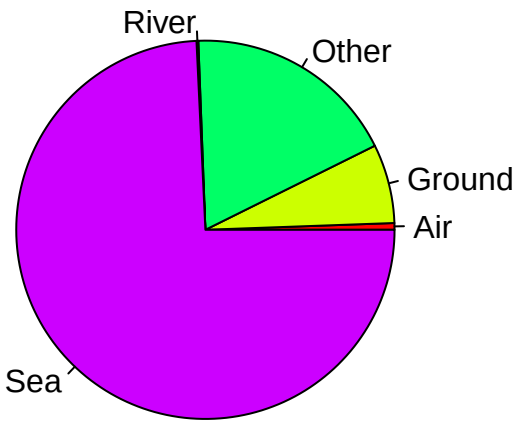

```
pie(table(data2$route[data2$product=='soybean_meal']), main = "Soybean meal routes", col = rainbow(leng
```



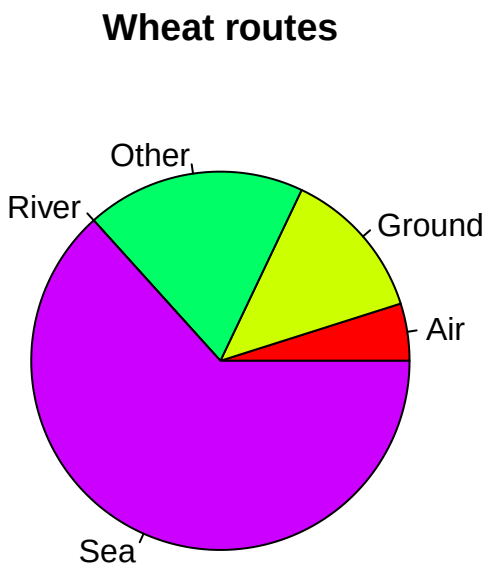

```
pie(table(data2$route[data2$product=='soybean_oil']), main = "Soybean oil routes", col = rainbow(length
```




```
pie(table(data2$route[data2$product=='sugar']), main = "Sugar routes", col = rainbow(length(cor)))
```




```
pie(table(data2$route[data2$product=='wheat']), main = "Wheat routes", col = rainbow(length(cor)))
```



With the exception of soybean oil, all products have their routes mostly by sea.

4)Which countries have been the most important trade partners for Brazil in terms of 'corn' and 'sugar' in the last 3 years?

```
rm(list = 'data2')
data2 = subset(data, data$product=='corn' )
data2$date = as.Date(data2$date)
data2= data2[ data2$date > as.Date("2017-12-01"), ]
temp = as.data.frame(table(data2$country)>50)
temp = subset(temp,temp==TRUE)

print(temp)
```

```
##                     table(data2$country) > 50
## Argentina                              TRUE
## Bangladesh                             TRUE
## Bolivia                                TRUE
## Colombia                               TRUE
## Dominican Republic                     TRUE
## Egypt                                  TRUE
## Iran                                   TRUE
## Japan                                  TRUE
## Malaysia                               TRUE
## Mexico                                 TRUE
## Morocco                                TRUE
## Paraguay                               TRUE
## Peru                                   TRUE
```

```
## Portugal                                  TRUE
## Saudi Arabia                              TRUE
## South Korea                               TRUE
## Spain                                     TRUE
## Taiwan                                    TRUE
## United States                             TRUE
## Vietnam                                   TRUE
```

Here we have a list of the trade partners that had at least 50 transations in terms of "corn" in the last 3 years

```
rm(list = 'data2')
data2 = subset(data, data$product=='sugar' )
data2$date = as.Date(data2$date)
data2= data2[ data2$date > as.Date("2017-12-01"), ]
temp = as.data.frame(table(data2$country)>100)
temp = subset(temp,temp==TRUE)

print(temp)
```

```
##                   table(data2$country) > 100
## Algeria                                TRUE
## Canada                                 TRUE
## China                                  TRUE
## Germany                                TRUE
## Hong Kong                              TRUE
## Liberia                                TRUE
## Malta                                  TRUE
## Marshall Islands                       TRUE
## Netherlands                            TRUE
## Norway                                 TRUE
## Panama                                 TRUE
## Saudi Arabia                           TRUE
## Singapore                              TRUE
## United States                          TRUE
## Uruguay                                TRUE
## Venezuela                              TRUE
## Yemen                                  TRUE
```

Here we have a list of the trade partners that had at least 100 transations in terms of "sugar" in the last 3 years

5)For each of the products in the dataset, show the 5 most important states in terms of exports?

```
rm(list = 'data2')
data2 = subset(data, data$type=='Export' )
sugar_state =sort( table(data2$state[data2$product=='sugar']),decreasing = TRUE)
print(sugar_state[1:5])
```

```
##
##    SP    GO    MG    PR    MS
## 18278  2732  2646  1993  1524
```

```
wheat_state =sort( table(data2$state[data2$product=='wheat']),decreasing = TRUE)
print(wheat_state[1:5])
```

```
##
##  PR  RS  SP  RO  RJ
```

```
## 308 194  80  27  25
```

```
soybeans_state =sort( table(data2$state[data2$product=='soybeans']),decreasing = TRUE)
print(soybeans_state[1:5])
```

```
##
##   MT   PR   GO   MS   MG
## 4980 3188 2166 1603 1472
```

```
soybean_oil_state =sort( table(data2$state[data2$product=='soybean_oil']),decreasing = TRUE)
print(soybean_oil_state[1:5])
```

```
##
##   PR   MT   SP   SC   RS
## 3476 2697 2131 2023 1420
```

```
soybean_meal_state =sort( table(data2$state[data2$product=='soybean_meal']),decreasing = TRUE)
print(soybean_meal_state[1:5])
```

```
##
##   MT   PR   RS   GO   MG
## 3953 3251 2657 1976 1602
```

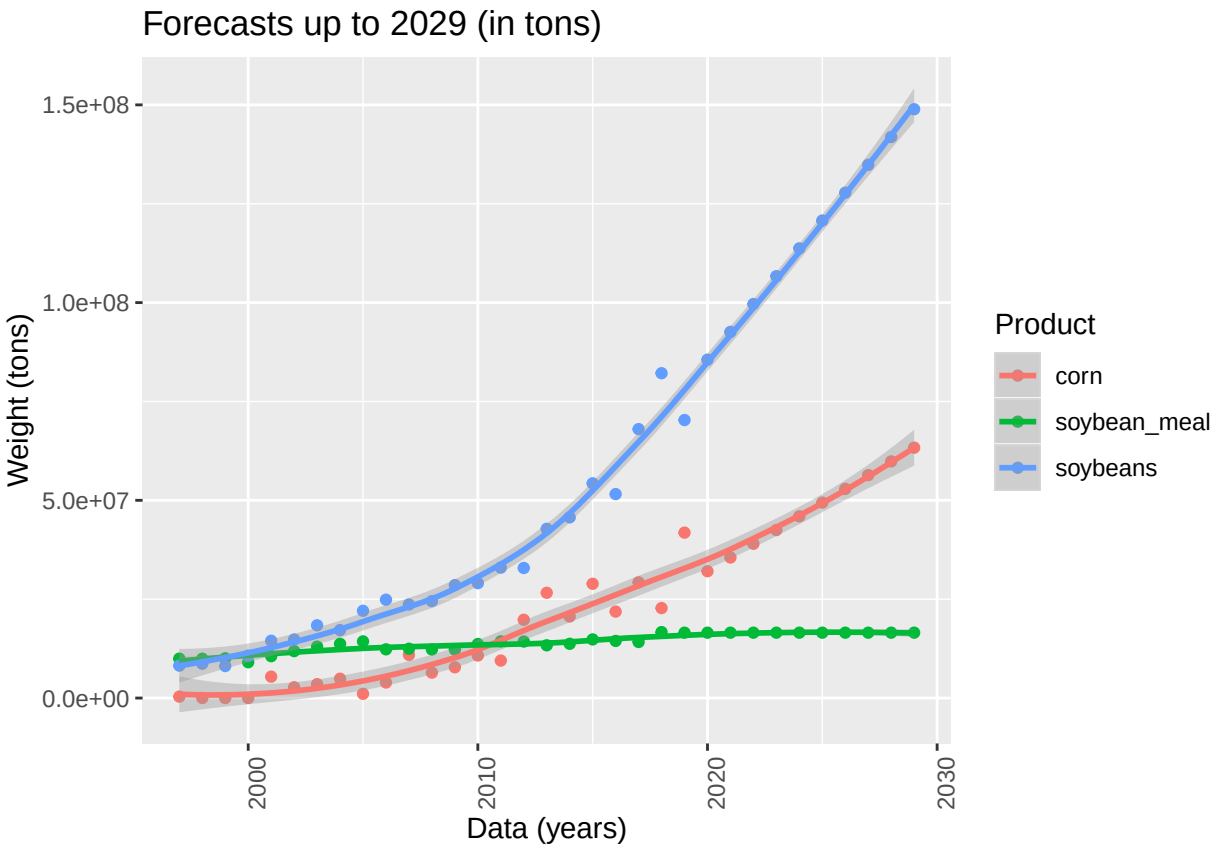
6)Now, we ask you to show your modelling skills. Feel free to use any type of modelling approach, but bear in mind that the modelling approach depends on the nature of your data, and so different models yield different estimates and forecasts. To help you out in this task we also provide you with a dataset of possible covariates (.xlsx). They all come from public sources (IMF, World Bank) and are presented in index number format. Question: What should be the total brazilian soybeans, soybean_meal, and corn export forecasts, in tons, for the next 11 years (2020-2030)? We're mostly interested in the annual forecast.

```
rm(list = 'data2', 'temp')
data2 = subset(data, data$type=='Export' )
data2 = subset(data2, data2$product=='soybeans' | data2$product=='soybean_meal' | data2$product=='corn'
data2$date = as.numeric(substr ( data2$date, 1,4 ))
year = matrix(2020:2030, 10,1)
```

```
## Warning in matrix(2020:2030, 10, 1): comprimento dos dados [11] não é um
## submúltiplo ou múltiplo do número de linhas [10]
```

```
#soybeans_forecast
temp = cbind(data2$date[data2$product=="soybeans"],data2$tons[data2$product=="soybeans"])
colnames(temp) = c("year","tons")
temp = as.data.frame(temp)
temp = group_by(temp,year)%>%summarise(tons=sum(tons))
temp2=as.data.frame(forecast(temp$tons, h=10))[1]
temp2=cbind(year,temp2)
dataSOYBEANS = as.data.frame(rbind(as.matrix(temp),as.matrix(temp2)))
product = matrix(rep('soybeans'),33,1)
dataSOYBEANS = cbind(dataSOYBEANS,product)
rownames(dataSOYBEANS) <- NULL

#soybean_meal_forecast
temp = cbind(data2$date[data2$product=="soybean_meal"],data2$tons[data2$product=="soybean_meal"])
colnames(temp) = c("year","tons")
temp = as.data.frame(temp)
temp = group_by(temp,year)%>%summarise(tons=sum(tons))
temp2=as.data.frame(forecast(temp$tons, h=10))[1]
```

```
temp2=cbind(year,temp2)
dataSOYBEAN_MEAL = as.data.frame(rbind(as.matrix(temp),as.matrix(temp2)))
product = matrix(rep('soybean_meal'),33,1)
dataSOYBEAN_MEAL = cbind(dataSOYBEAN_MEAL,product)
rownames(dataSOYBEAN_MEAL) <- NULL

#corn_forecast
temp = cbind(data2$date[data2$product=="corn"],data2$tons[data2$product=="corn"])
colnames(temp) = c("year","tons")
temp = as.data.frame(temp)
temp = group_by(temp,year)%>%summarise(tons=sum(tons))
temp2=as.data.frame(forecast(temp$tons, h=10))[1]
temp2=cbind(year,temp2)
dataCORN = as.data.frame(rbind(as.matrix(temp),as.matrix(temp2)))
product = matrix(rep('corn'),33,1)
dataCORN = cbind(dataCORN,product)
rownames(dataCORN) <- NULL

#Binding and plotting
dataFinal=as.data.frame(rbind(dataCORN,dataSOYBEAN_MEAL,dataSOYBEANS))

p <- ggplot(dataFinal, aes(x = dataFinal$year, y = dataFinal$tons , color = as.factor(dataFinal$product)
  geom_point() +
  geom_smooth(method = "loess")+
  theme(axis.text.x = element_text(angle = 90))+
  labs(title = "Forecasts up to 2029 (in tons)", x = "Data (years)", y = "Weight (tons)", color = "Produ

p
```



Projecting the data with the 'forecast' library; in 2029 we can expect to export 1.489081e + 08 ton of Soybeans, 1.653039e+07 ton of Soybean meal and 6.331813e+07 ton of Corn.