Untitled

Required libraries

```
library(RPostgreSQL)
```

```
## Loading required package: DBI
```

```
library(tidyverse)
```

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

```
## ✓ ggplot2 2.2.1     ✓ purrr   0.2.4
## ✓ tibble  1.4.1     ✓ dplyr   0.7.4
## ✓ tidyr   0.7.2     ✓ stringr 1.2.0
## ✓ readr   1.1.1     ✓ forcats 0.2.0
```

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

```
library(dbplyr)
```

```
##
## Attaching package: 'dbplyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     ident, sql
```

```
library(rjson)
library(DBI)
library(lubridate)
```

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

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

```
library(tibble)
library(olsrr)
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
library(ggplot2)
library(ggExtra)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

Get the auxiliary data

```
source("get_HMRC_aux_data.R")
list1 <- get_HMRC_aux_data()
comcode <- data.frame(Reduce(rbind, list1[1]))
port    <- data.frame(Reduce(rbind, list1[2]))
country <- data.frame(Reduce(rbind, list1[3]))
write.csv(comcode,file="comcode.csv")
write.csv(country,file="country.csv")
```

—Partners codes

Poland: 616

Spain : 724

Brazil: 76

UK : 826

—Three commodities codes

Chicken: 02071

Cucumber: 070700

Beef: 160250

GET COMTRADE DATA

```
source("get_Comtrade_data.R")
#Define commodity and partner country
com_id    <- "070700"
partner_id <- 724
#Comtrade SQL request
stime <- Sys.time()
df1   <- get_Comtrade_data(201401,201601,"default",com_id,as.character(partner_id))
etime <- Sys.time()
print(etime-stime)
```

```
## Time difference of 4.210923 mins
```

Tidy Comtrade data

```
#Group by commodity code for the same good if necessary (different cuts for chicken...)
df2  <- df1 %>% group_by(period,trade_flow,reporter,reporter_code,partner,partner_code) %>%
               summarize(net_weight_kg   = sum(netweight_kg),
               trade_value_usd = sum(trade_value_usd)) %>% ungroup()
#Compute the price in usd per kg
df2  <- df2  %>% mutate(price_usd_kg = trade_value_usd/net_weight_kg)
#Turn period into a proper date
df2  <- df2 %>% mutate(period_date = ymd(paste(period,"01",sep="")))
#Remove missing observations
df2 <- df2[complete.cases(df2),]
#Get the comtrade data for imports into the uk for the given commodity
comtrade_imports_into_uk <- df2 %>%
  filter(reporter=="United Kingdom") %>%
  filter(trade_flow=="Imports")
```

GET HMRC DATA

```
source("get_HMRC_data.R")
stime <- Sys.time()
HMRC_EU_import_food_data <- get_HMRC_data(arrivals)
```

```
## [1] "Medium cuppa?"
```

```
etime <- Sys.time()
print(etime-stime)
```

```
## Time difference of 7.174511 mins
```

```
(col_names               <- t(as.data.frame(colnames(HMRC_EU_import_food_data))))
```

```
##                                  [,1]           [,2]
## colnames(HMRC_EU_import_food_data) "smk_comcode" "smk_record_type"
##                                  [,3]           [,4]
## colnames(HMRC_EU_import_food_data) "smk_cod_seq" "smk_cod_alpha"
##                                  [,5]             [,6]
## colnames(HMRC_EU_import_food_data) "smk_trade_ind" "smk_coo_seq"
##                                  [,7]
## colnames(HMRC_EU_import_food_data) "smk_coo_alpha"
##                                  [,8]
## colnames(HMRC_EU_import_food_data) "smk_nature_of_transaction"
##                                  [,9]
## colnames(HMRC_EU_import_food_data) "smk_mode_of_transport"
##                                  [,10]
## colnames(HMRC_EU_import_food_data) "smk_period_reference"
##                                  [,11]                 [,12]
## colnames(HMRC_EU_import_food_data) "smk_suite_indicator" "smk_sitc"
##                                  [,13]
## colnames(HMRC_EU_import_food_data) "smk_ip_comcode"
```

```
##                                    [,14]
## colnames(HMRC_EU_import_food_data) "smk_no_of_consignments"
##                                    [,15]             [,16]
## colnames(HMRC_EU_import_food_data) "smk_stat_value" "smk_nett_mass"
##                                    [,17]
## colnames(HMRC_EU_import_food_data) "smk_supp_unit"
```

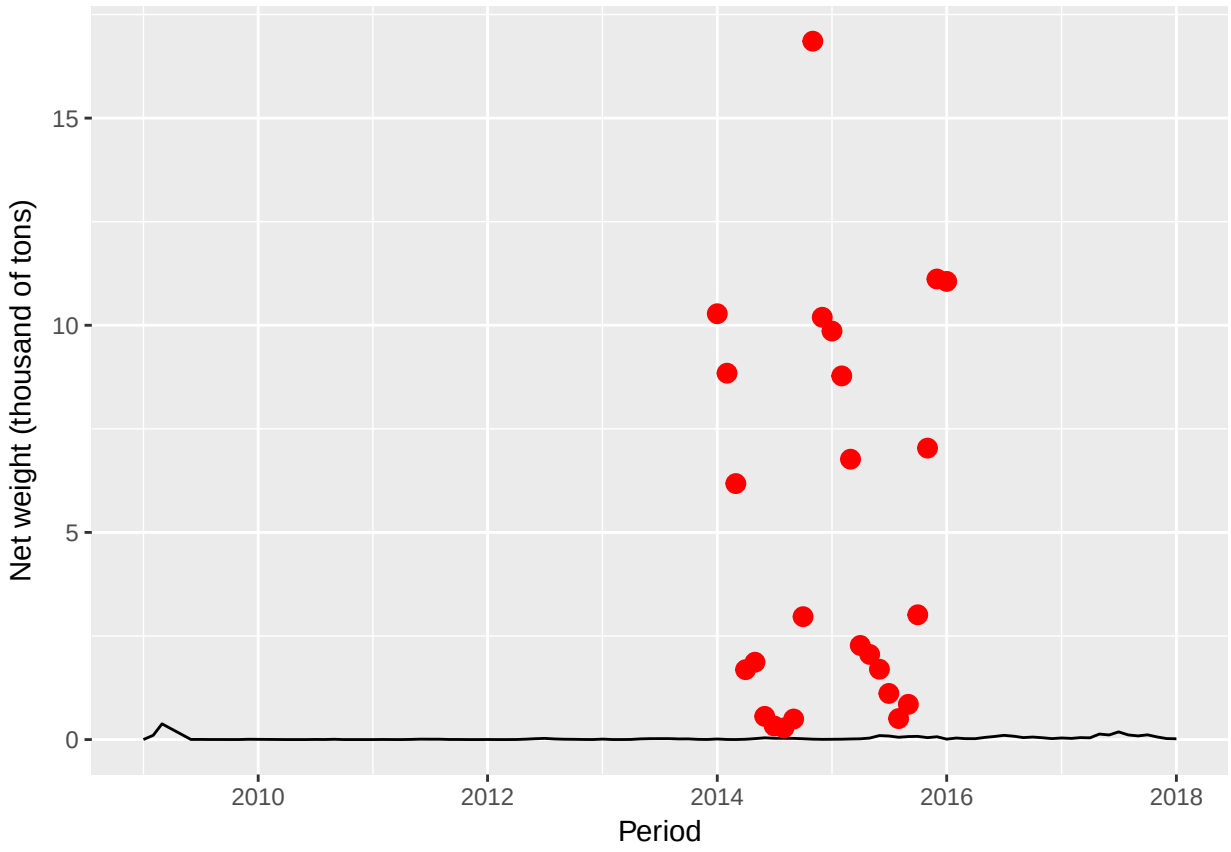
```
#Filter the data for the selected commodity_code
tmp  <- HMRC_EU_import_food_data
#tmp1 <- tmp %>% filter(str_sub(smk_comcode,1,str_length(com_id)) == com_id)
tmp2 <- tmp[str_detect(tmp$smk_comcode,paste('^',com_id,sep='')),]
#Remove crazy year
current_year <- 2018
tmp2 <- tmp2 %>% filter(as.numeric(smk_period_reference)<100*(current_year+1))
#Ignore some variables
tmp2 <- tmp2 %>%
select(-smk_coo_seq,-smk_coo_alpha) %>%
select(-smk_nature_of_transaction,-smk_mode_of_transport,-smk_no_of_consignments) %>%
select(-smk_suite_indicator,-smk_sitc,-smk_ip_comcode) %>% select(-smk_supp_unit,-smk_trade_ind,-smk_re
#Rename variables
tmp2 <- tmp2 %>% rename(commodity_code  = "smk_comcode")
tmp2 <- tmp2 %>% rename(partner_code    = "smk_cod_seq")
tmp2 <- tmp2 %>% rename(partner_id      = "smk_cod_alpha")
tmp2 <- tmp2 %>% rename(period          = "smk_period_reference")
tmp2 <- tmp2 %>% rename(trade_value_spd = "smk_stat_value")
tmp2 <- tmp2 %>% rename(netweight_kg    = "smk_nett_mass")
#Sterling pounds to US dollars
tmp2 <- tmp2 %>% mutate(trade_value_usd = trade_value_spd * 1.41) %>% select(-trade_value_spd)
```

```
#Group by commodity code for the same good if necessary (different cuts for chicken...)
tmp3 <- tmp2 %>% group_by(period,partner_id,partner_code) %>%
                 summarize(net_weight_kg   = sum(netweight_kg),
                 trade_value_usd = sum(trade_value_usd)) %>% ungroup()
#Compute the price in usd per kg
tmp3 <- tmp3 %>% mutate(price_usd_kg = trade_value_usd/net_weight_kg)
#Turn period into a proper date
tmp3 <- tmp3 %>% mutate(period_date = ymd(paste(as.character(as.numeric(period)),"01",sep="")))
```

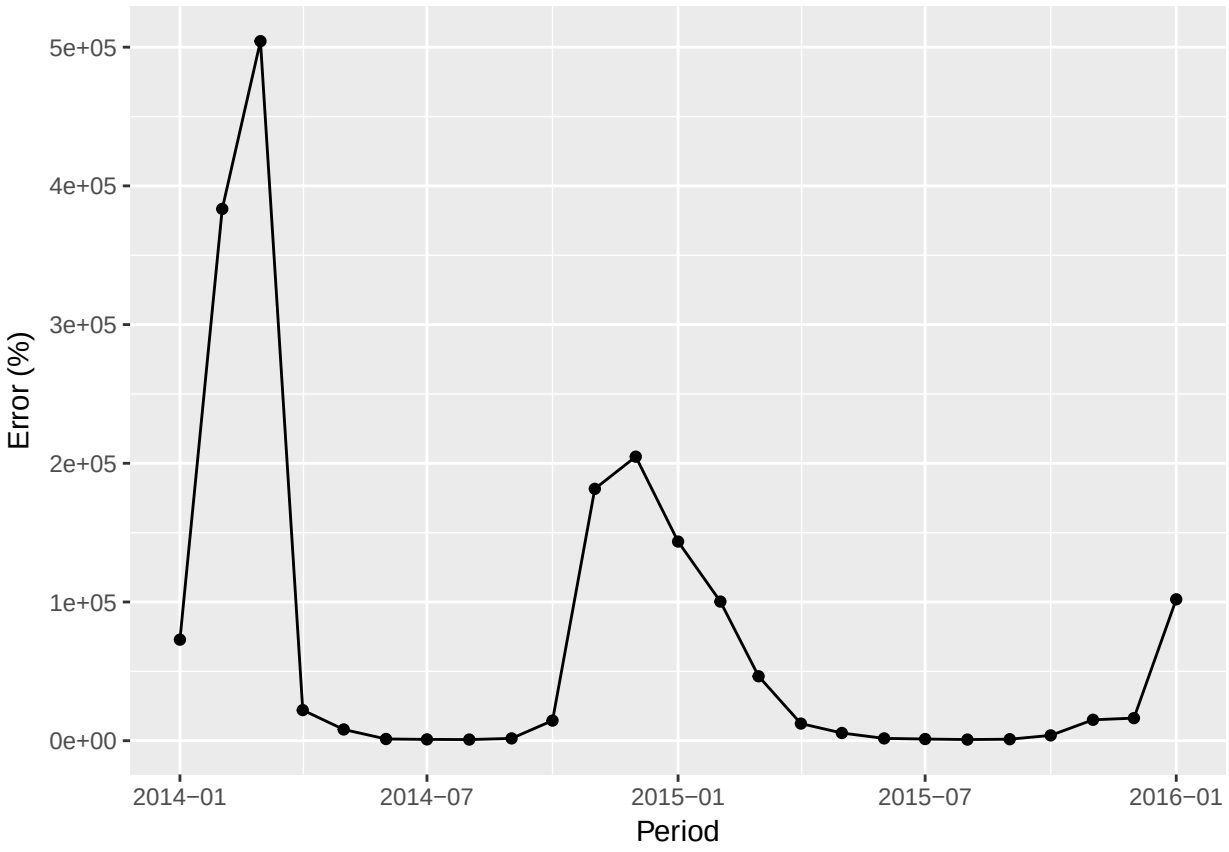
```
## Warning: 29 failed to parse.
```

```
#Remove missing observations
tmp4 <- tmp3[complete.cases(tmp3),]
tmp5 <- tmp4 %>% filter(trade_value_usd > 0 & net_weight_kg > 0)
#Get the comtrade data for imports into the uk for the given commodity
HMRC_imports_into_uk <- tmp5 %>% filter(partner_id == "PL")
```

```
ggplot(NULL) + geom_line(data=HMRC_imports_into_uk,
                         mapping = aes(x=period_date,y=net_weight_kg/1e6)) +
               geom_point(data=comtrade_imports_into_uk,
                         mapping = aes(x=period_date,y=net_weight_kg/1e6),color="red",size=3) +
  labs(x="Period",y="Net weight (thousand of tons)")
```




```
comb <- inner_join(comtrade_imports_into_uk,HMRC_imports_into_uk,by="period_date")
weight <- comb %>% select(starts_with("net_weight_kg"),period_date) %>%
  mutate(error = (100*(net_weight_kg.x-net_weight_kg.y)/net_weight_kg.y))
ggplot(data=weight,aes(x=period_date)) + geom_line(aes(y=error)) +
                                        geom_point(aes(y=error)) +
  labs(x="Period",y="Error (%)")
```


5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
Error (%)
2014−01
2014−07
2015−01
2015−07
2016−01
Period