Spanish cucumber imports into the UK

Required libraries

```
require("RPostgreSQL")
```

```
## Loading required package: RPostgreSQL
```

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

```
library(ggplot2)
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 1.2.1 ──
```

```
## ✓ tibble  1.4.2     ✓ purrr   0.2.4
## ✓ tidyr   0.8.0     ✓ stringr 1.3.0
## ✓ readr   1.1.1     ✓ forcats 0.3.0
```

```
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
```

Get the HMRC 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]))
```

Find cucumber comcodes - Thank you Alex for making life nice and easy for us!

```
cc_all_cucumber <- comcode[grep('CUCUMBER',  toupper(comcode$description)),]
```

Get HMRC import data from EU (arrivals df)

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

## [1] "Medium cuppa?"

(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"
```

Select Spain cucumber import info

```
HMRC_columns  <- col_names[c(1,4,10,14,15,16)]
HMRC_country  <- "ES"
HMRC_comcode  <- "07070005"
model_data    <- HMRC_EU_import_food_data %>%
                 select(HMRC_columns) %>%
filter(smk_cod_alpha == HMRC_country & smk_comcode ==
HMRC_comcode) %>%
                 select(-smk_cod_alpha, -smk_comcode)
```

Clean data: remove any Na values, 0 values and date values == “0000000”

```
model_data_clean <- model_data %>% na.omit %>%
                    filter(smk_nett_mass!=0 & smk_no_of_consignments != 0 &
smk_stat_value != 0) %>%
filter(smk_period_reference!="0000000") %>%
mutate(abs(smk_nett_mass) & abs(smk_no_of_consignments) &
abs(smk_stat_value)) %>%
                    unique()
```

Convert period column to date format

```
model_data_clean <- model_data_clean %>%
                    mutate(smk_period_reference =
substr(smk_period_reference,                                    4,7)) %>%
                    mutate(smk_period_reference =
paste0(smk_period_reference,"01")) %>%
                    mutate(smk_period_reference =
as.Date(smk_period_reference,
                                                              "%y%m%d"))
```

Get the cos/kg ($/kg)

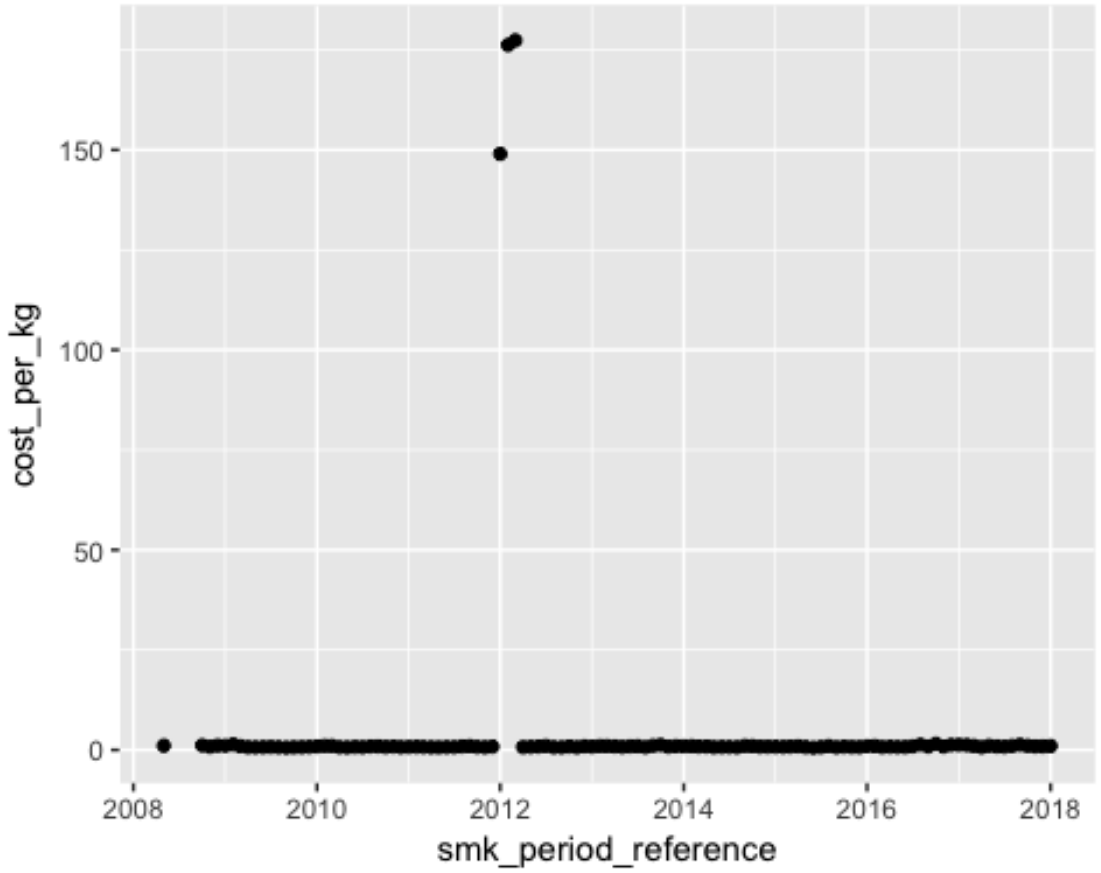
```
model_data_clean <- model_data_clean %>%
                    mutate(cost_per_kg = smk_stat_value / smk_nett_mass)
```

Aggregate data into periods of months

```
model_data_group <- model_data_clean %>% group_by(smk_period_reference) %>%
summarise(smk_nett_mass = sum(smk_nett_mass),
smk_no_of_consignments  = sum(smk_no_of_consignments),
smk_stat_value          = sum(smk_stat_value),
                    cost_per_kg             = mean(cost_per_kg))
```

What does the cost/kg of the cucumber imported into the UK look like over time?

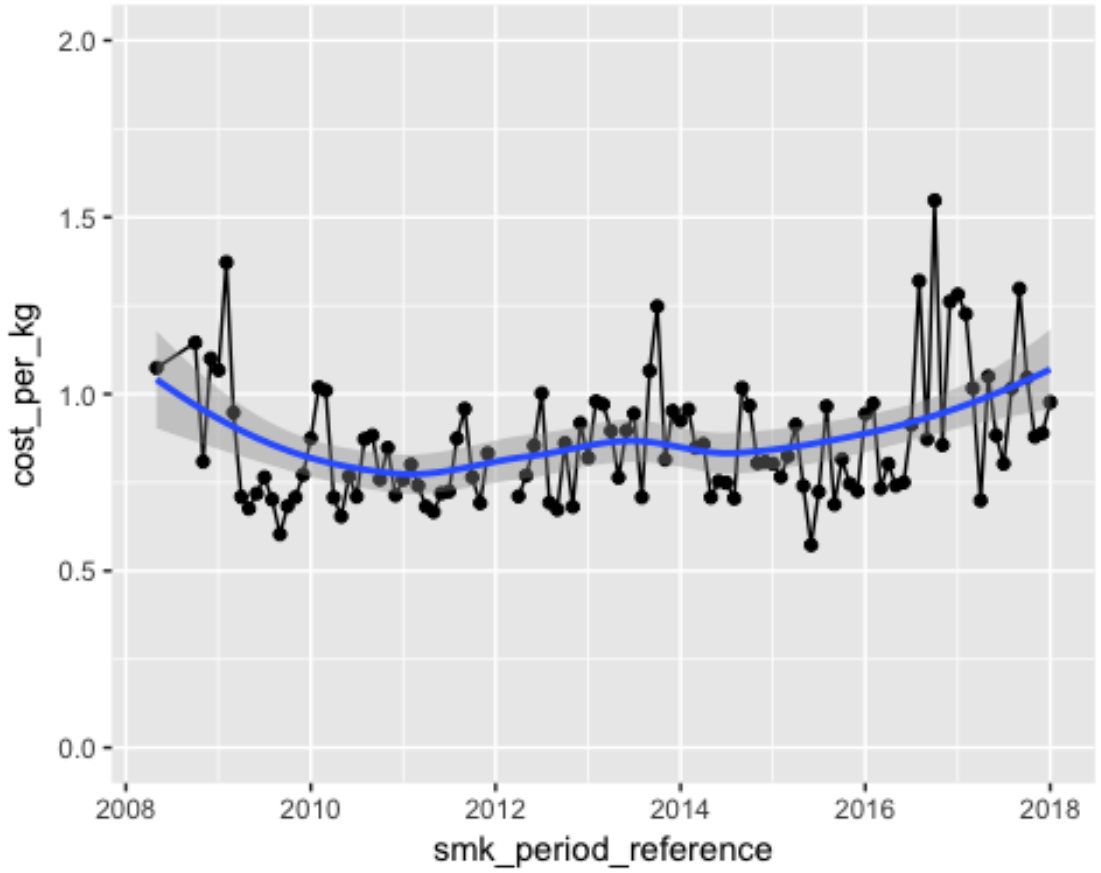
```
ggplot(model_data_group, aes(smk_period_reference, cost_per_kg)) +
        geom_point()
```



Very high cost/kg during 2012/2013 - $150/kg seems to be a bit too exorbitant for cucumber! What does the cost/kg look like without the points above $150/kg?

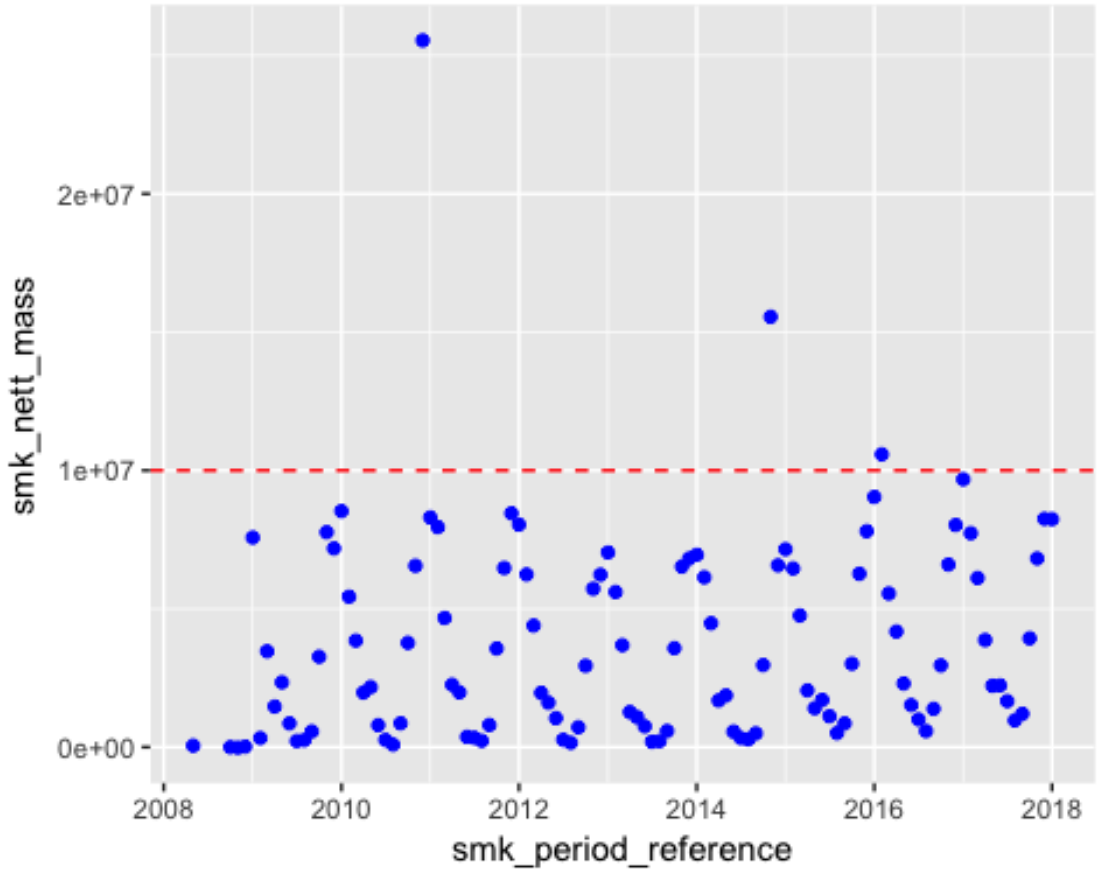
```
ggplot(model_data_group, aes(smk_period_reference, cost_per_kg)) +
  geom_point() + geom_line() + ylim(0,2) + geom_smooth()
```

```
## `geom_smooth()` using method = 'loess'
```



What does the nett mass of the cucumber imported into the UK look like over time?

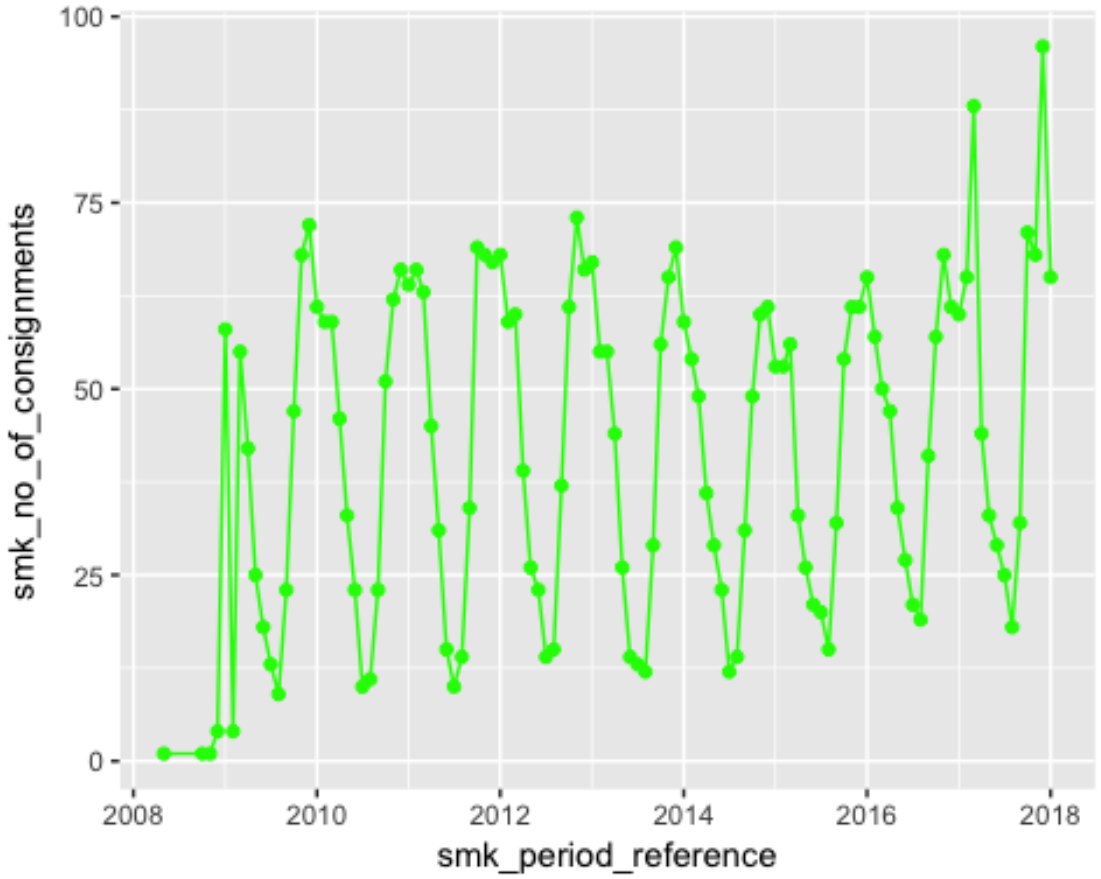
```
ggplot(model_data_group, aes(smk_period_reference, smk_nett_mass)) +
        geom_point(colour = "blue") +
        geom_hline(yintercept=1E7, linetype="dashed", color = "red")
```



- Imports of cucumber appears to be seasonal and follow an oscillating pattern with exception for several points (Occur Nov - Jan) where there are spikes in 2010 & 2014 and possibly 2009, 2016 & 2017.
- Also looking at cost/kg over time, note how cost/kg simmilar in months before and/or after at some of these points in time.

Did the number of consignments also increase during the spikes?

```
ggplot(model_data_group, aes(smk_period_reference, smk_no_of_consignments)) +
        geom_line(colour = "green") +
        geom_point(colour = "green")
```



Conclusions and questions

- Possible flag for irregular increased mass cucumbers imported into the UK
- Possible flag for prices being simmilar before and/or after month of increased mass imported
- Where do these mysterious cucumbers come from? Comtrade data may help here.
- To flag future anomalies can the oscillating and seasonal nett mass of imports of cucumbers from Spain into the UK be modelled?
- Janis any ideas with modelling continuous data over time - maybe try simple non-linear function using X=period,no. consignments,cost/kg and Y = nett mass cucumbers and try regression? Other model ideas?

- Looking ahead, can links be made to determine the source of these extra cucumbers? Possible starting points to to model import/export network with focus on mass balances?