**Q. Import data sets HH\_SAMPLE\_POWER\_CONSUMPTION\_GRID \_A.csv into R Studio**

**Download the zip file from moodle Topic “Visualisation”  “Examples” and extract to a local folder**

**library(readr)**

**consumption <- read\_delim("C:/PATH\_TO\_FILE/HH\_SAMPLE\_POWER\_CONSU MPTION\_GRID\_A.csv", ";", na = "empty", trim\_ws = TRUE)**

1. library(readr)
2. consumption<-read\_delim("C:/Users/Nikhila Mariam Biju/Desktop/datamining/New folder/HH\_SAMPLE\_POWER\_CONSUMPTION\_GRID\_A.csv",";",na="empty",trim\_ws = TRUE)

Parsed with column specification:

cols(

GRID\_IDENT = col\_character(),

MONTH\_BILLING = col\_character(),

SWITCHING\_REASON = col\_character(),

DECLARATION\_CONSUMP\_CUST = col\_double(),

DECLARATION\_CONSUMP\_GRID = col\_double(),

CORRECTIVE\_VALUE = col\_double(),

DAYS\_INVOICED = col\_double(),

CONSUMPTION\_INVOICED = col\_double(),

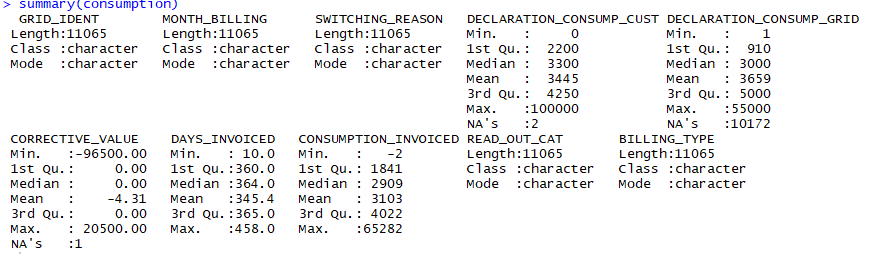
READ\_OUT\_CAT = col\_character(),

BILLING\_TYPE = col\_character()

)

1. summary(consumption)

- Gives an overall summary about the columns of the consumption table with length, class, mode and other information. A screenshot for the one is shown below



**Q. Create a new table or tibble with the name hh\_spc: – month\_billing = MONTH\_BILLING – swt\_reason = SWITCHING\_REASON – cust\_decl\_kwh = DECLARATION\_CONSUMP\_CUST – grid\_decl\_kwh = DECLARATION\_CONSUMP\_CUST + CORRECTIVE\_VALUE – bill = BILLING\_TYPE – inv\_kwh\_365 = CONSUMPTION\_INVOICED \* 365 / DAYS\_INVOICED • Use round function to cut decimals**

1. library(tibble) :

Used to load all the packages related to the data frame.

1. hh\_spc<-tibble( month\_billing = consumption$MONTH\_BILLING,swt\_reason = consumption$SWITCHING\_REASON,cust\_decl\_kwh=consumption$DECLARATION\_CONSUMP\_CUST,grid\_decl\_kwh=consumption$DECLARATION\_CONSUMP\_GRID+consumption$CORRECTIVE\_VALUE,bill=consumption$BILLING\_TYPE,inv\_kwh\_365=consumption$CONSUMPTION\_INVOICED\*365/consumption$DAYS\_INVOICED)

**Q.** **Limit the selection to: – inv\_kwh\_365 < 10000 – cust\_decl\_kwh < 7500 – cust\_decl\_kwh > 0**

1. library(dplyr):

This library is used to provide filter and other data manipulation operations that be used with the table hh\_spc.

1. hh\_spc<-filter(hh\_spc,inv\_kwh\_365<10000,cust\_decl\_kwh<7500,cust\_decl\_kwh > 0)

* It returns modified data sets depending on the conditions that is applied to the data sets.
* Here, filter function when applied to table hh\_spc returns the results of colums inv\_kwh with the values less than 10000, cust\_decl with a value greater than 0 and lesser than 7500

**Order by month\_billing**

1. hh\_spc <- arrange(hh\_spc, month\_billing)

Arrange Function here arranges the table hh\_spc in the order sequence of month\_billing

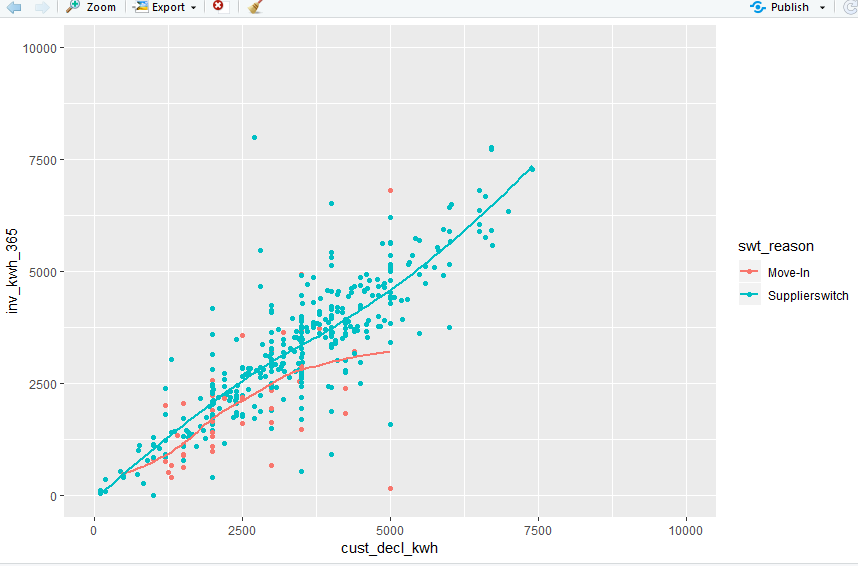
**Q. Plot a sample of 400 of these point & including smoothing function – Compare customers’ declared consumption and (invoiced) 365 day consumption – Separate switching reason by color – Use se = FALSE for the smoothing**

1. library(ggplot2)

Loads the packages that are required for plotting

1. ggplot(data = sample\_n(tbl=hh\_spc, size=400),aes(x = cust\_decl\_kwh,y = inv\_kwh\_365,color = swt\_reason)) + geom\_point() + geom\_smooth(se = FALSE) + xlim(0,10000) + ylim(0,10000)

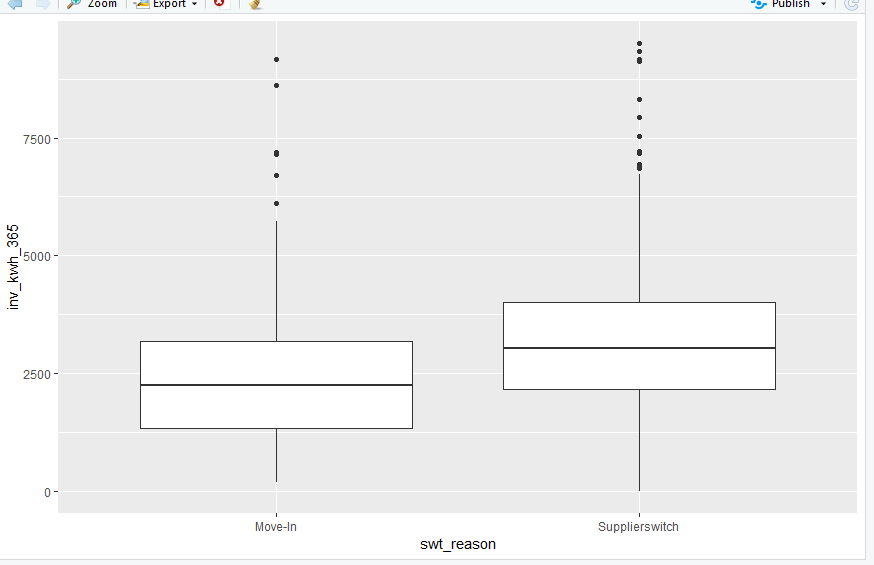
* geom\_smooth() smooth the plots over the points against cust\_decl\_kwh and inv\_kwh\_365
* xlim, ylim are the cartesion coordinate points



* Observation: People who conduct a supplierswitch can estimate their power demand much more precise than people who move in
* Because each time it can viewed that the variation of power demand in move-in occurs frequently

**Q. Create a boxplot with switching reason and 365 day consumption by using a 10% sample. (Hint: If it does not work change x and y attributes)**

1. ggplot(data = sample\_frac(tbl=hh\_spc, size=0.1),aes(x = swt\_reason, y = inv\_kwh\_365)) +geom\_boxplot()



* Observation: People who conduct a supplierswitch use more power than people who have just moved in