Datavisualisation automatique de grand jeux de données

R User Group

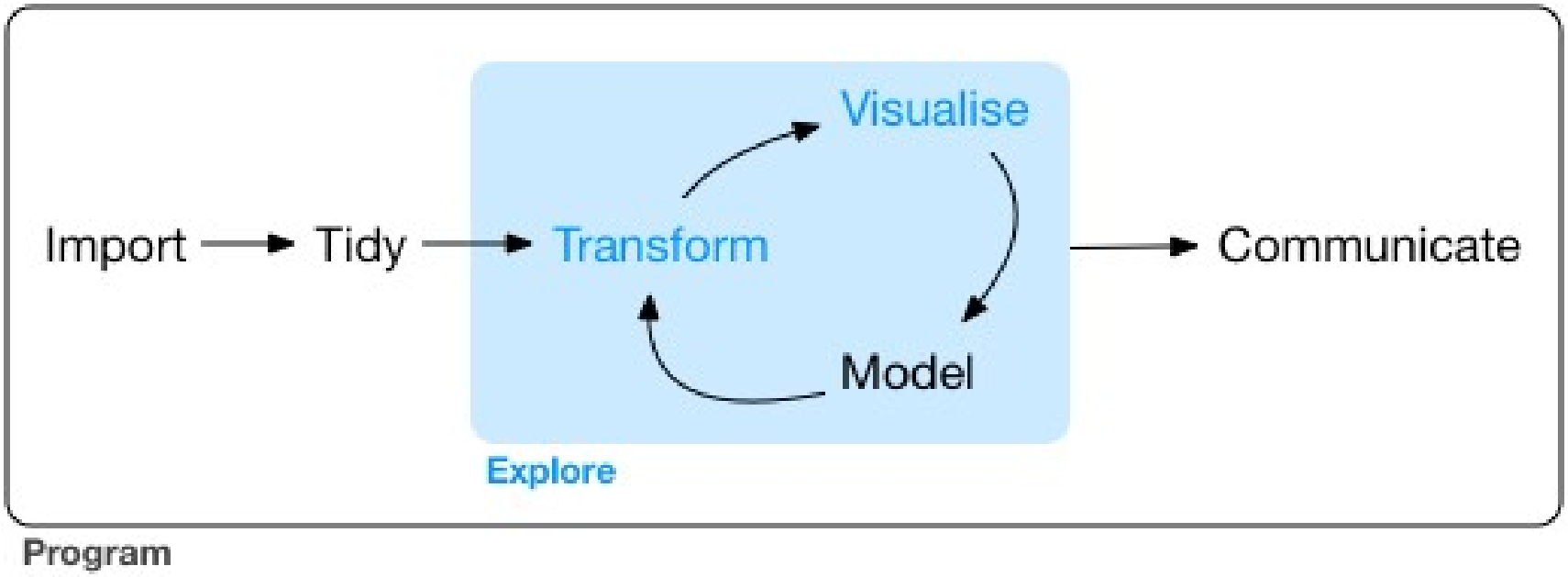
Toulouse

Disclaimer:

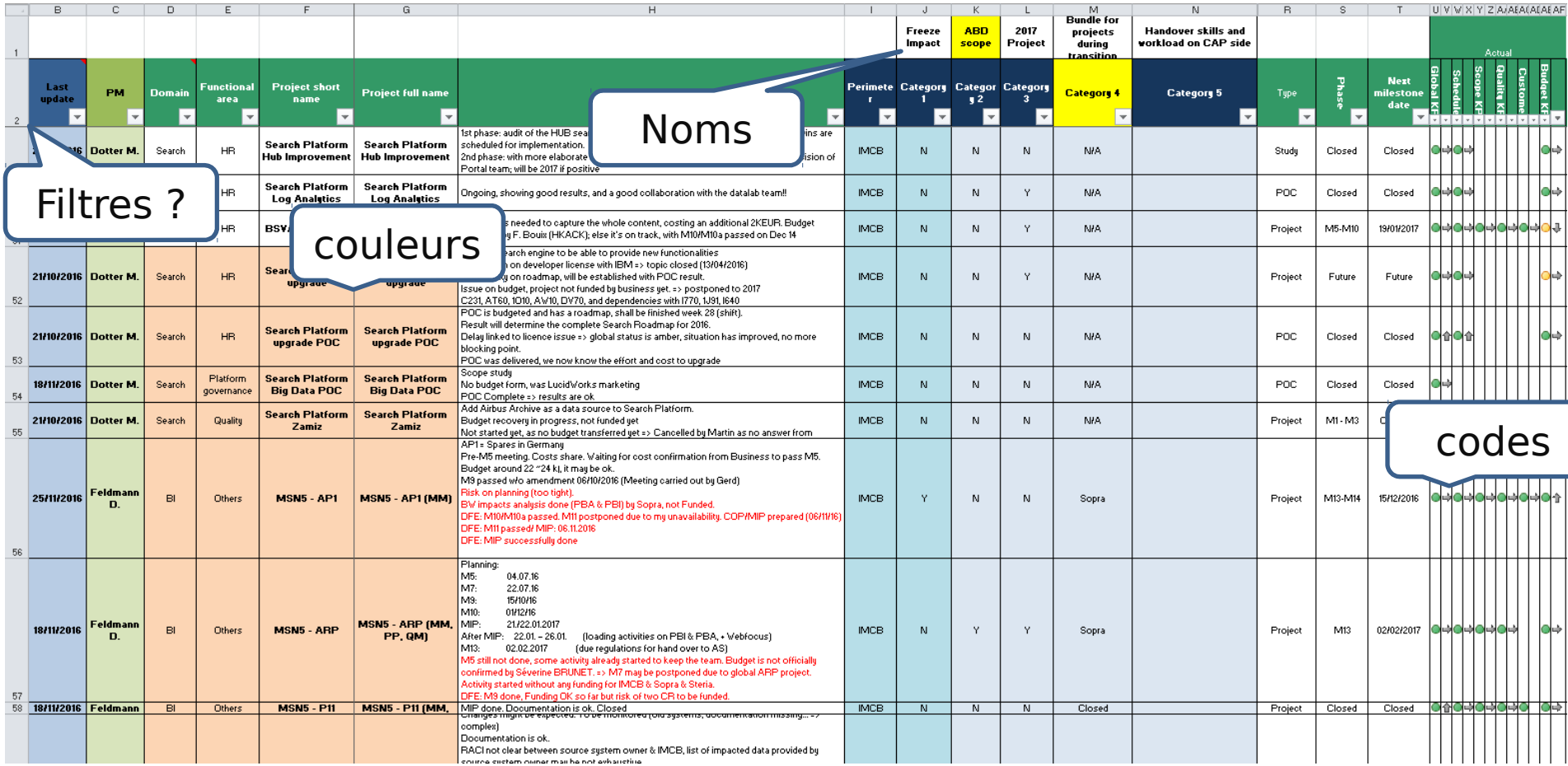
I am not representing my employer AIRBUS in this talk

I cannot confirm nor deny if AIRBUS is using any of the methods, tools
results etc. mentioned in this talk

La data-visualisation pour explorer les données



Un grand Dataset ? (1) MS-Excel !



	B	C	D	E	F	G	H	I	J	K	L	M	N	R	S	T	U V W X Y Z AA AB AC AD AE AF
1									Freeze Impact	ABD scope	2017 Project	Bundle for projects during transition	Handover skills and workload on CAP side				Actual
2	Last update	PM	Domain	Functional area	Project short name	Project full name		Perimeter	Category 1	Category 2	Category 3	Category 4	Category 5	Type	Phase	Next milestone date	Global KPI / Schedule KPI / Scope KPI / Quality KPI / Custome / Budget KPI
	16	Dotter M.	Search	HR	Search Platform Hub Improvement	Search Platform Hub Improvement	1st phase: audit of the HUB sea ins are scheduled for implementation. 2nd phase: with more elaborate ision of Portal team; will be 2017 if positive	IMCB	N	N	N	N/A		Study	Closed	Closed	
				HR	Search Platform Log Analytics	Search Platform Log Analytics	Ongoing, showing good results, and a good collaboration with the datalab team!!	IMCB	N	N	Y	N/A		POC	Closed	Closed	
				HR	BSV		s needed to capture the whole content, costing an additional 2KEUR. Budget y F. Bouix (HKACK); else it's on track, with M10/M10a passed on Dec 14	IMCB	N	N	Y	N/A		Project	M5-M10	19/01/2017	
52	21/10/2016	Dotter M.	Search	HR	Search upgrade	upgrade	earch engine to be able to provide new functionalities on developer license with IBM => topic closed (13/04/2016) y on roadmap, will be established with POC result. Issue on budget, project not funded by business yet. => postponed to 2017 C231, AT60, 1010, AW10, DV70, and dependencies with I770, IJ91, I640	IMCB	N	N	Y	N/A		Project	Future	Future	
53	21/10/2016	Dotter M.	Search	HR	Search Platform upgrade POC	Search Platform upgrade POC	POC is budgeted and has a roadmap, shall be finished week 28 (shift). Result will determine the complete Search Roadmap for 2016. Delay linked to licence issue => global status is amber, situation has improved, no more blocking point. POC was delivered, we now know the effort and cost to upgrade	IMCB	N	N	N	N/A		POC	Closed	Closed	
54	18/11/2016	Dotter M.	Search	Platform governance	Search Platform Big Data POC	Search Platform Big Data POC	Scope study No budget form, was LucidWorks marketing POC Complete => results are ok	IMCB	N	N	N	N/A		POC	Closed	Closed	
55	21/10/2016	Dotter M.	Search	Quality	Search Platform Zamiz	Search Platform Zamiz	Add Airbus Archive as a data source to Search Platform. Budget recovery in progress, not funded yet Not started yet, as no budget transferred yet => Cancelled by Martin as no answer from	IMCB	N	N	N	N/A		Project	M1 - M3	C	
56	25/11/2016	Feldmann D.	BI	Others	MSN5 - AP1	MSN5 - AP1 (MM)	AP1 = Spares in Germany Pre-M5 meeting. Costs share. Waiting for cost confirmation from Business to pass M5. Budget around 22 ~24 k), it may be ok. M9 passed w/o amendment 06/10/2016 (Meeting carried out by Gerd) Risk on planning (too tight). BW impacts analysis done (PBA & PBI) by Sopra, not Funded. DFE: M10/M10a passed. M11 postponed due to my unavailability. COP/MIP prepared (06/11/16) DFE: M11 passed/ MIP: 06.11.2016 DFE: MIP successfully done	IMCB	Y	N	N	Sopra		Project	M13-M14	15/12/2016	
57	18/11/2016	Feldmann D.	BI	Others	MSN5 - ARP	MSN5 - ARP (MM, PP, QM)	Planning: M5: 04.07.16 M7: 22.07.16 M9: 15/10/16 M10: 01/12/16 MIP: 21/22.01.2017 After MIP: 22.01. – 26.01. (loading activities on PBI & PBA, + Webfocus) M13: 02.02.2017 (due regulations for hand over to AS) M5 still not done, some activity already started to keep the team. Budget is not officially confirmed by Séverine BRUNET. => M7 may be postponed due to global ARP project. Activity started without any funding for IMCB & Sopra & Steria. DFE: M9 done, Funding OK so far but risk of two CR to be funded.	IMCB	N	Y	Y	Sopra		Project	M13	02/02/2017	
58	18/11/2016	Feldmann	BI	Others	MSN5 - P11	MSN5 - P11 (MM,	MIP done. Documentation is ok. Closed Changes might be expected. To be monitored (old systems, documentation missing... => complex) Documentation is ok. RACI not clear between source system owner & IMCB, list of impacted data provided by source system owner may be not exhaustive	IMCB	N	N	N	Closed		Project	Closed	Closed	

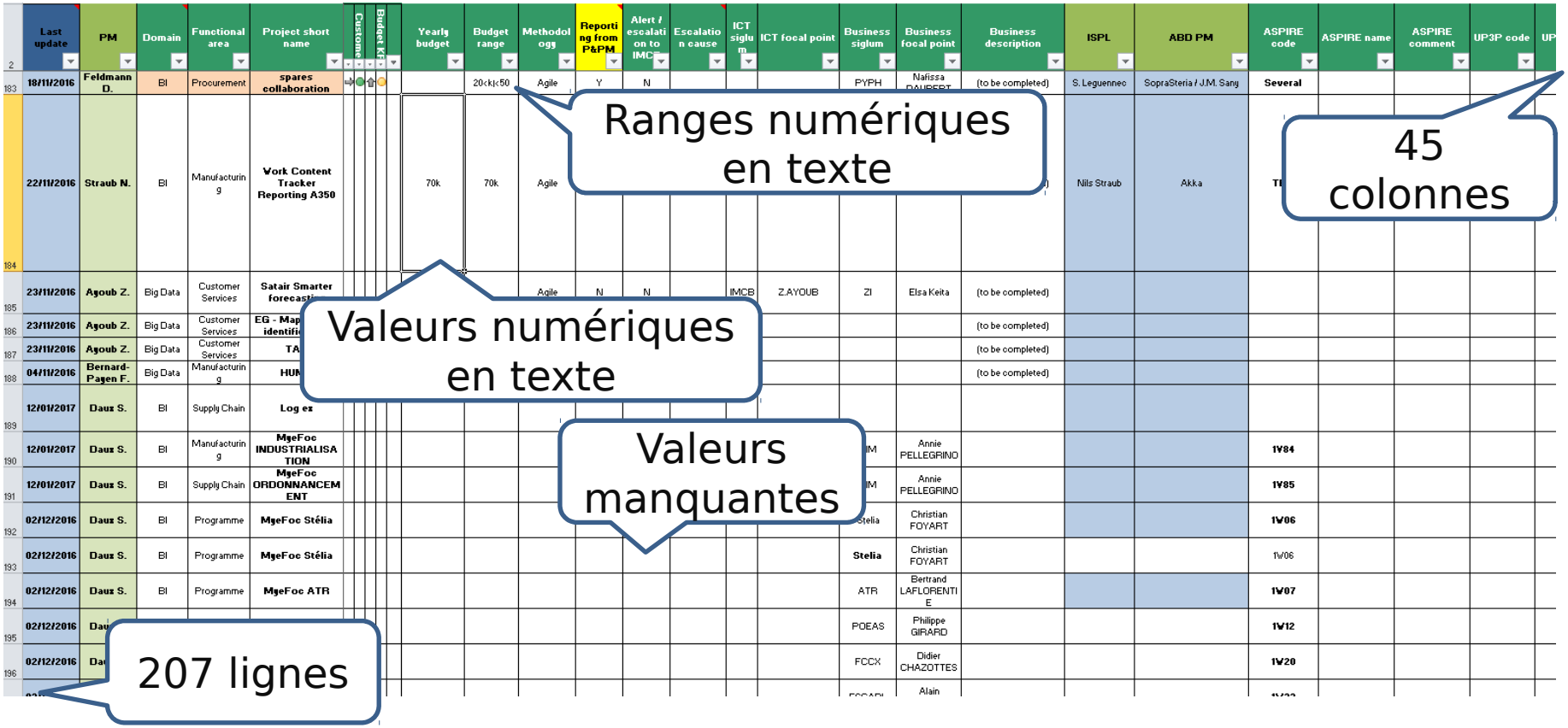
Noms

Filtres ?

couleurs

codes

Un grand Dataset ? (1) MS-Excel !



Un grand Dataset ? (1) R summary()

```
> summary(rag_df)
  Update.week       Last.update                  
 Min.    : 2.00   Min.    :2016-03-10 00:00:00   Ler
 1st Qu.:46.00   1st Qu.:2016-11-18 00:00:00   Cla
 Median :47.00   Median :2016-11-23 00:00:00   Moc
 Mean   :42.94   Mean   :2016-11-25 09:00:52  
 3rd Qu.:48.00   3rd Qu.:2016-12-05 00:00:00  
 Max.   :51.00   Max.   :2017-01-13 00:00:00  
 NA's   :2                                    
 Project.short.name Project.full.name    Comments

   ...

        UP3P.code                                    UP3P.name  
 -          :  1   None                               :  9  
 ?          :  1   BW Reporting Evolution             :  4  
 2211       :  1   AG-2598 - HR Analytics Spotfire    :  2  
 2585       :  1   ?                                  :  1  
 Ai-2345    :  1   AG-2403 - Foundation Wave 1 2015 BI:  1  
 To be done:  1   (Other)                            : 25  
 NA's       :199   NA's                               :163  
```

argh

oups

whizz

crac

Un grand Dataset ? (1) R summary()

Import → Tidy → Transform

We import the data

```{r, include=FALSE}
library(readxl)
library(plotluck)
library(dplyr)
library(stringr)
filepath<-"C:\\temp\\"
filelist<-list.files(path=filepath,pattern="*RAG*.xlsx")[[1]]
rag_df <- read_excel(paste0(filepath,filelist),sheet = "RAG", skip = 1)
rag_df <- tibble::set_tidy_names(rag_df,syntactic =TRUE)
```

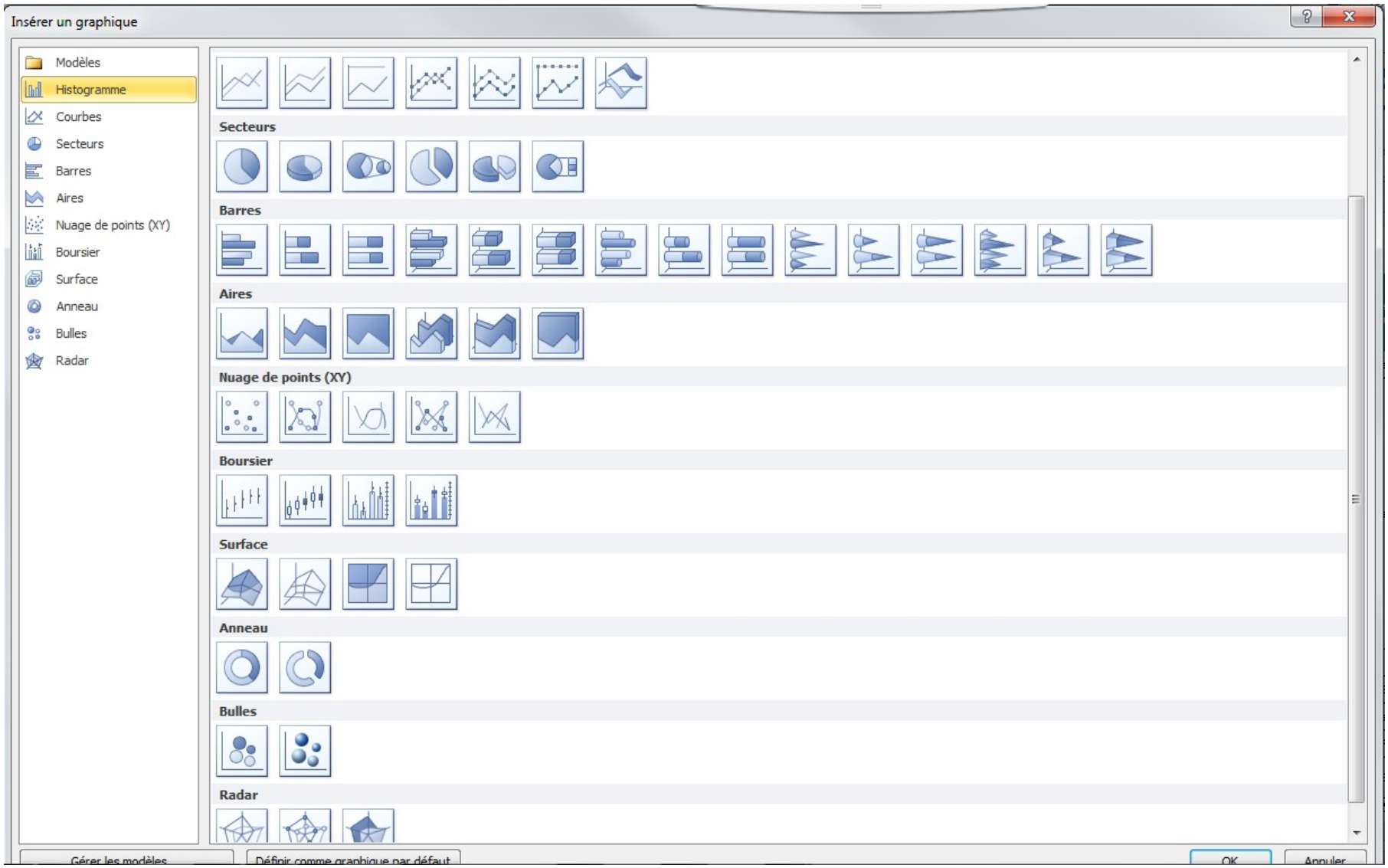
Let's clean the data

```{r}
#turn Yearly.budget into value
rag_df$Yearly.budget <- as.numeric(rag_df$Yearly.budget)
#turn "TBD"" and "-" ... values into NA for any column being character
chr_columns<-purrr::map_lgl(rag_df,is.character) %>% as.vector
fromto=c("tbd"=NA_character_,"-"=NA_character_,"none"=NA_character_,
         "n/a"=NA_character_,"^na$"=NA_character_,"to be defined"=NA_character_,
         "to be done"=NA_character_,"\\?"=NA_character_,"\\?\\?\\?"=NA_character_,
         "\\(to be completed\\)"=NA_character_)
rag_df[,chr_columns] <- lapply(rag_df[,chr_columns], function(col)
  str_replace_all(str_to_lower(col), pattern=fromto )) %>%
  as.data.frame %>%
  mutate_if(is.character,as.factor)
#explicit column names
names(rag_df)[10:13]<-c("freeze_impact","abd_scope","2017","bundler")
```

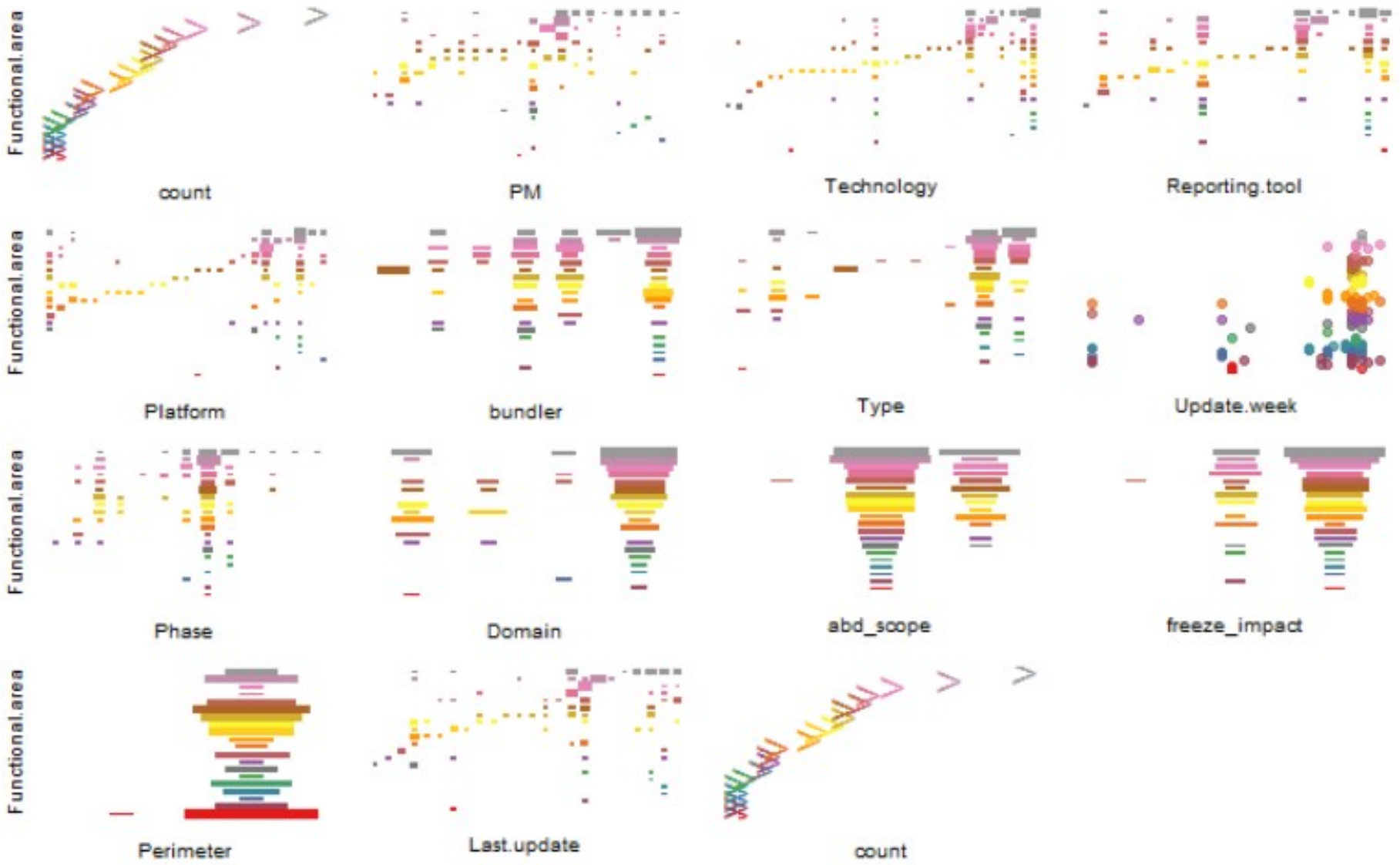
```
:select(-matches("Category\\.5"),-matches(".*omments.*"),-matches(".*\\.name"))
```

```
odate                         PM              Domain        Functional.area Perimeter  fr
 2016-03-10 00:00:00   D[illegible] S.  :50   Application:  5   Manufacturing:36   AH  :  1   N
 2016-11-18 00:00:00   V[illegible] S.  :32   BI          :158   Programme    :27   IMCB:204   Y
 2016-11-23 00:00:00   F[illegible] D.  :22   Big Data    : 32   Procurement  :20              NA
 2016-11-25 09:00:52   A[illegible] Z.  :15   Search      : 10   Supply Chain :17
 2016-12-05 00:00:00   L[illegible] B.  :10                      Quality      :16
 2017-01-13 00:00:00   B[illegible] M.  : 9                      (Other)      :84
                       (Other)          :67                      NA's         : 5
nology  Reporting.tool                  Type            Phase     Next.milestone.date   Globa
  :46   BW      :48     Project          :109   Closed    : 29   January   :28       Min.
s:32    Focus   :41     CR               : 53   Continuous: 25   Closed    :26       1st Qu.
  :21   BI4     :38     POC              : 20   M10a-M11  : 18   Continuous:25       Median
  :18   Hadoop  :20     Platform Governance:  9   Future    : 15   Future    :14       Mean
  :18   Webfocus:18     Study            :  6   N/A       : 10   N/A       :10       3rd Qu.
  :63   (Other) :28     (Other)          :  7   (Other)   :102   (Other)   :94       Max.
  : 7   NA's    :12     NA's             :  1   NA's      :  6   NA's      : 8       NA's
KPI      Scope.trend    Quality.KPI    Quality.trend  Customer.satisfaction.KPI Customer.Sati
0.000   Min.   :1.00   Min.   :1.000   Min.   :1.00   Min.   :0.000             Min.   :1.000
2.000   1st Qu.:1.00   1st Qu.:2.000   1st Qu.:1.00   1st Qu.:2.000             1st Qu.:1.000
2.000   Median :1.00   Median :2.000   Median :1.00   Median :2.000             Median :1.000
1.903   Mean   :1.06   Mean   :1.963   Mean   :1.03   Mean   :1.936             Mean   :1.049
2.000   3rd Qu.:1.00   3rd Qu.:2.000   3rd Qu.:1.00   3rd Qu.:2.000             3rd Qu.:1.000
2.000   Max.   :2.00   Max.   :2.000   Max.   :2.00   Max.   :2.000             Max.   :2.000
81      NA's   :89     NA's   :96      NA's   :105    NA's   :111               NA's   :124
dget.range     Methodology Reporting.from.P.PM Alert...escalation.to.IMCB
0  : 36   GPP       :79   n   :  1   N   :163   Budget Issue: transf
0  : 25   Agile     :42   N   :123   Y   : 10   Delay of project
200: 18   Agile like:10   Y   : 50   NA's: 32   Problems of resource
00 :  6   Others    : 7   NA's: 31              Resources
500:  3   Project   : 6                         Technical issues wit
  :  5   (Other)   :17                         NA's
```

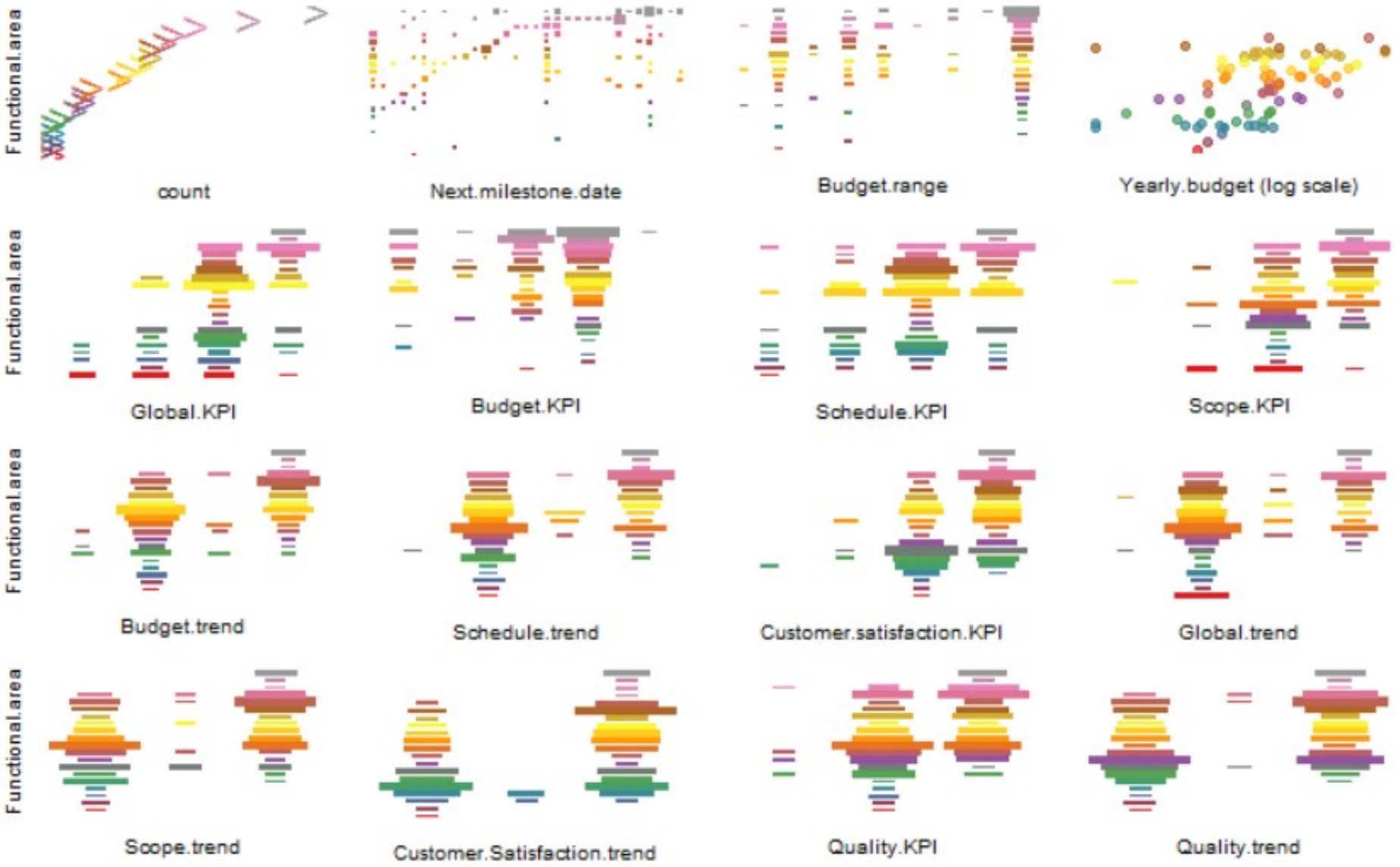
Un grand Dataset (1) quelle visualisation ?



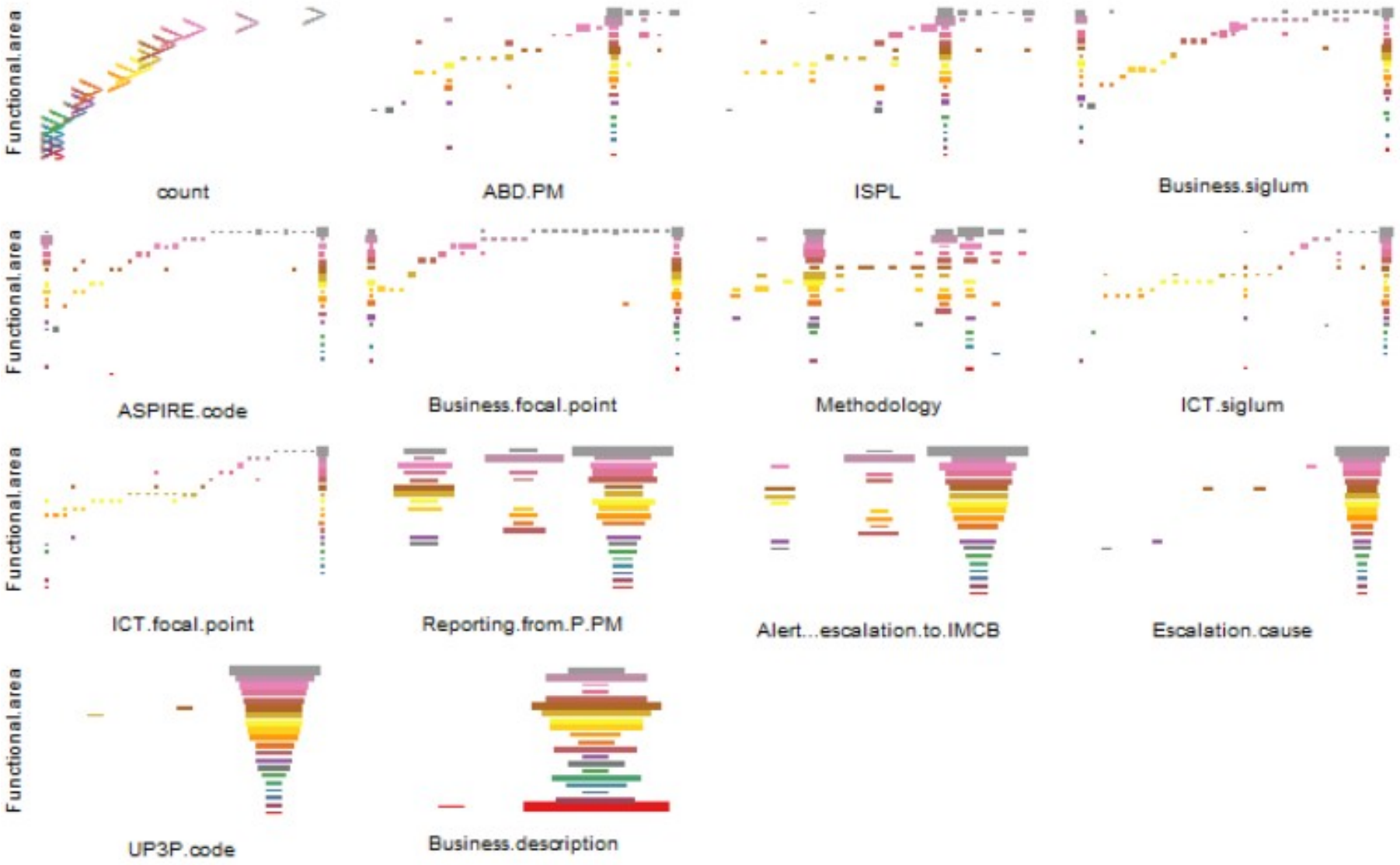
Un grand Dataset ? (1) R plotluck()



Un grand Dataset ? (1) R plotluck()



Un grand Dataset ? (1) R plotluck()



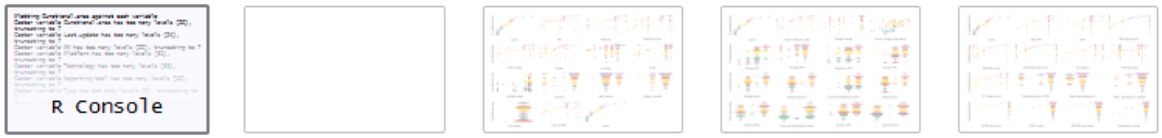
Un grand Dataset ? (1) R plotluck()

Analyse data sizing and quality (missing values, skewness, structure, summaries,)

```
```{r single-file exploratory graph}

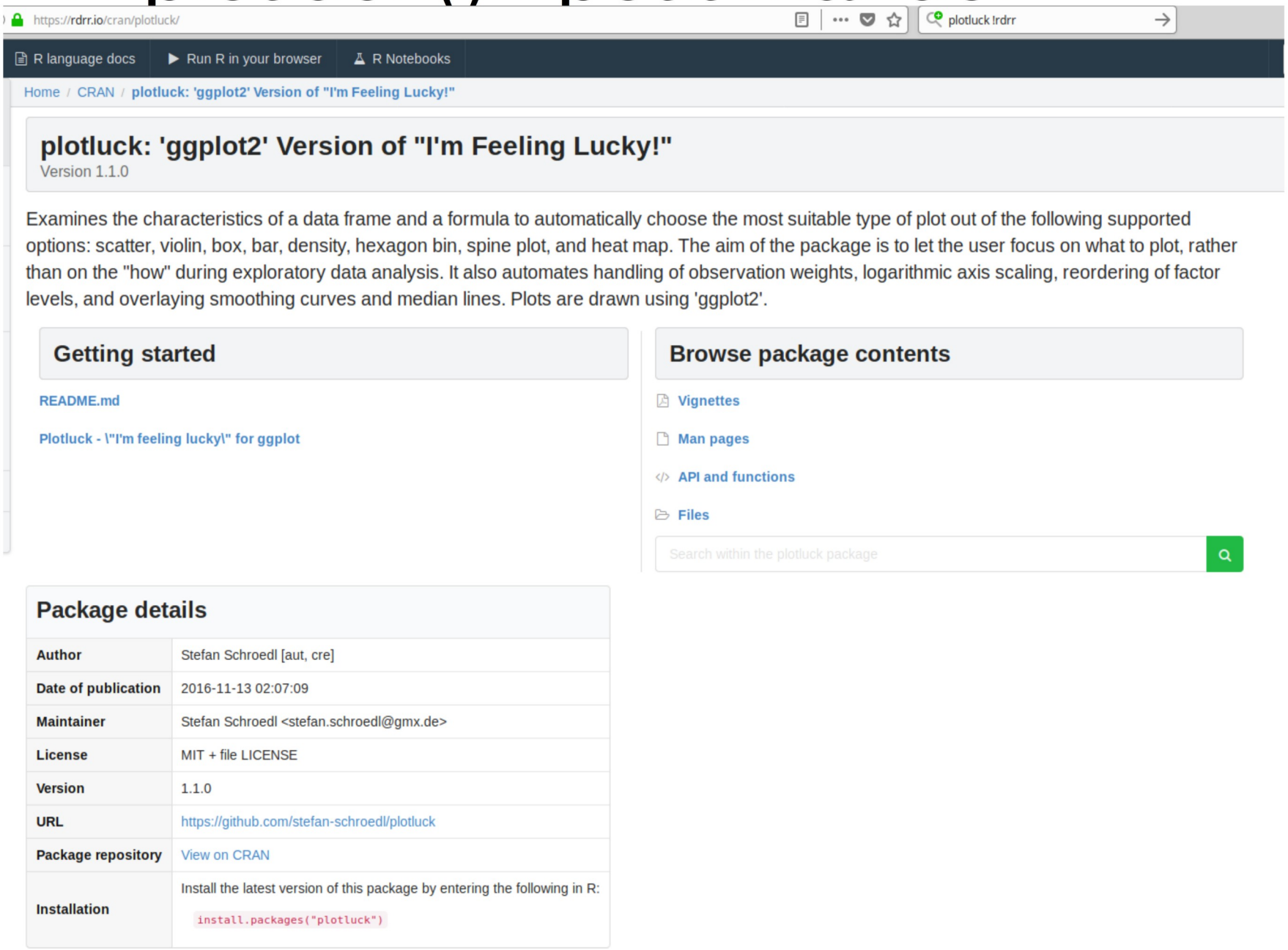
#rag_df %>%  select(-c(1,3,5,26,35,46)) %>% plotluck(formula=rag_project~.,opts= plotluck.options(verbose=TRUE))
rag_df %>% select(1:15) %>% plotluck(formula=Functional.area~.,opts= plotluck.options(verbose=TRUE))
rag_df %>% select(5,16:30) %>% plotluck(formula=Functional.area~.,opts= plotluck.options(verbose=TRUE))
rag_df %>% select(5,31:45) %>% plotluck(formula=Functional.area~.,opts= plotluck.options(verbose=TRUE))

```
```




```
Plotting Functional.area against each variable
Factor variable Functional.area has too many levels (20), truncating to 7
Factor variable Last.update has too many levels (24), truncating to 7
Factor variable PM has too many levels (22), truncating to 7
Factor variable Platform has too many levels (31), truncating to 7
Factor variable Technology has too many levels (35), truncating to 7
Factor variable Reporting.tool has too many levels (19), truncating to 7
Factor variable Type has too many levels (9), truncating to 7
Factor variable Phase has too many levels (46), truncating to 7
Ordering variables according to conditional entropy:
             var cond.ent
 Functional.area 0.000000
              PM 1.738159
        Platform 2.005648
  Reporting.tool 2.032351
      Technology 2.037585
           Phase 2.131956
         bundler 2.240711
            Type 2.255924
     Update.week 2.422087
          Domain 2.576475
       abd_scope 2.709201
   freeze_impact 2.730187
       Perimeter 2.775422
     Last.update 2.787745
```

R plotluck() : petit manuel



R plotluck() : petit manuel

```
1   plotluck(data, formula, weights, opts = plotluck.options(), ...)
```

Arguments

`data`	a data frame.
`formula`	an object of class `formula`: a symbolic description of the relationship of up to three variables.

Formula	Meaning	Plot types
`y~1`	Distribution of single variable	Density, histogram, scatter, dot, bar
`y~x`	One explanatory variable	Scatter, hex, violin, box, spine, heat
`y~x+z`	Two explanatory variables	heat, spine
`y~1|z` or `y~x|z`	One conditional variable	Represented through coloring or facetting
`y~1|x+z`	Two conditional variables	Represented through facetting

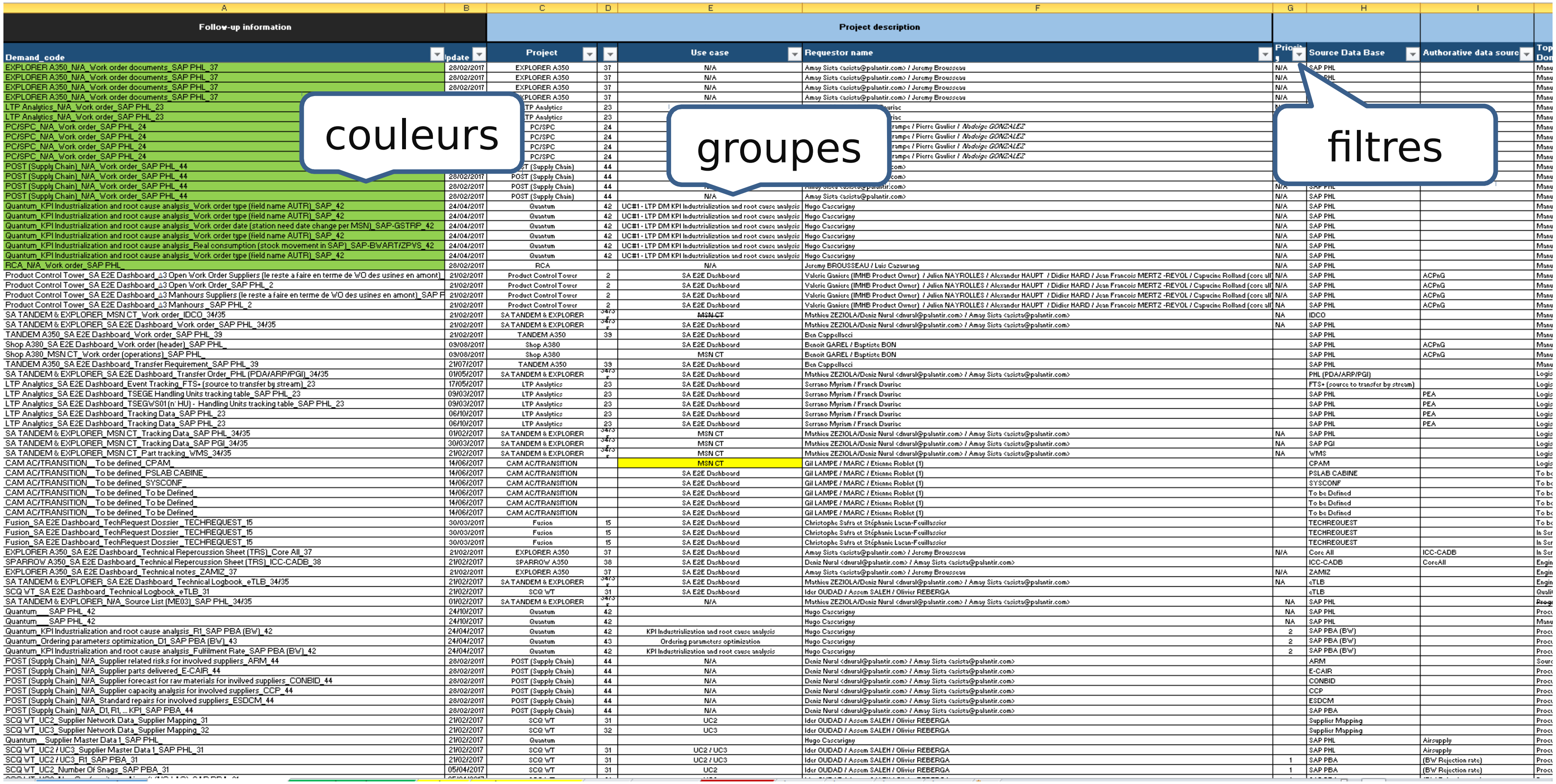
In addition to these base plot types, the dot symbol `"."` can also be used, and denotes all variables in the data frame. This gives rise to a lattice or series of plots (use with caution, can be slow).

Formula	Meaning
`.~1`	Distribution of each variable in the data frame, separately
`y~.`	Plot `y` against each variable in the data frame
`.~x`	Plot each variable in the data frame against `x`
`.~.`	Plot each variable in the data frame against each other.

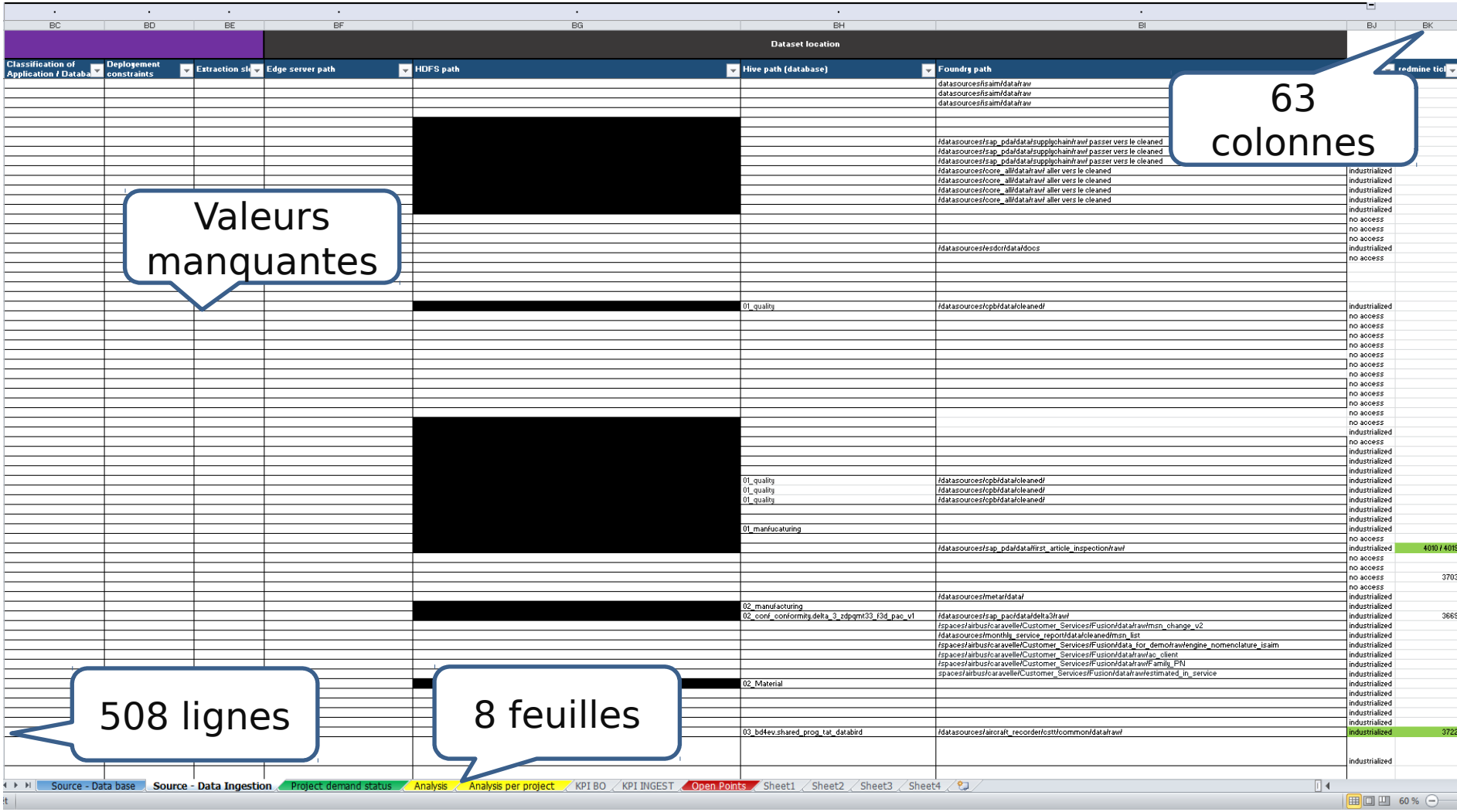
See also section "Generating multiple plots at once" below.

`weights`	observation weights or frequencies (optional).
`opts`	a named list of options (optional); See also `plotluck.options`.

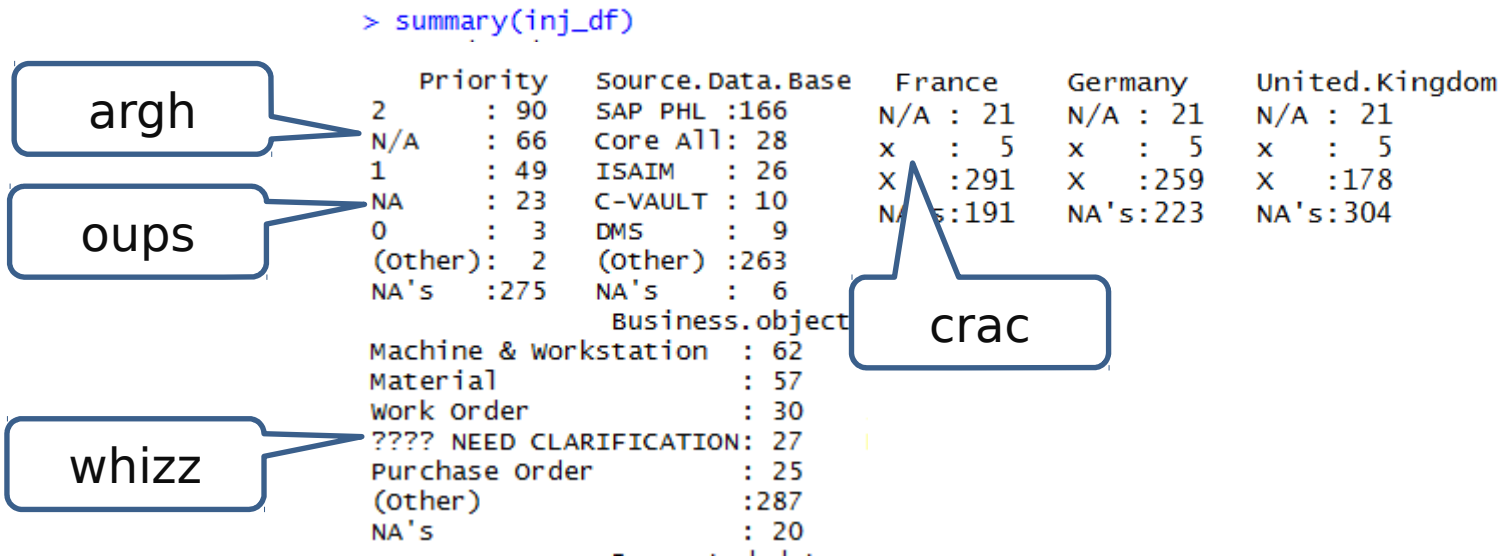
Un grand Dataset ? (2) MS-Excel !



Un grand Dataset ? (2) MS-Excel !



Un grand Dataset ? (2) R summary()



Un grand Dataset ? (2) R summary()

Import → Tidy → Transform

We import the Data

```{r, include=FALSE}
library(readxl)
library(plotluck)
library(dplyr)
library(tibble)
library(stringr)
inj_df <- read_excel("C:/Temp/Caravelle_data_ingestion_status.xlsx",
                     sheet = "Source - Data Ingestion", skip = 1,trim_ws = T)
inj_df<-set_tidy_names(inj_df,syntactic =TRUE)
```

Let's clean the data

```{r}
#turn "TBD"" and "-" ... values into NA for column 3+
chr_columns<-purrr::map_lgl(inj_df,is.character) %>% as.vector
fromto=c("tbd"=NA_character_,"-"=NA_character_,"none"=NA_character_,
         "n/a"=NA_character_,"^na$"=NA_character_,"to be defined"=NA_character_,
         "To be done"=NA_character_,"\\?"=NA_character_,"\\?\\?\\?"=NA_character_,
         "\\?\\?\\?\\? need clarification"=NA_character_)
inj_df[,chr_columns] <- lapply(inj_df[,chr_columns], function(col)
  str_replace_all(str_to_lower(col), pattern=fromto )) %>%
  as.data.frame %>%
  mutate_if(is.character,as.factor)
```

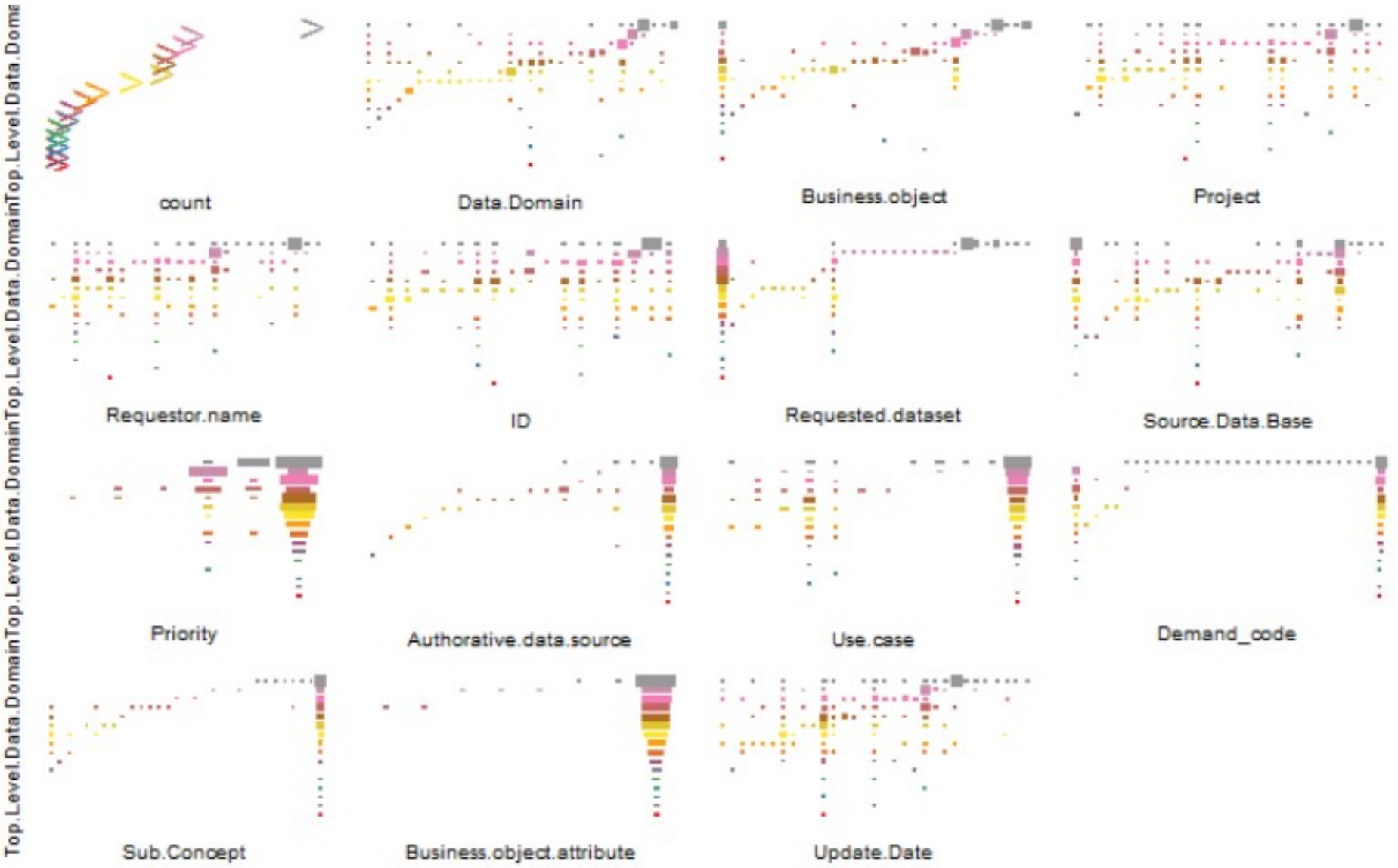
Un grand Dataset ? (2) R summary()

Let's see the summary

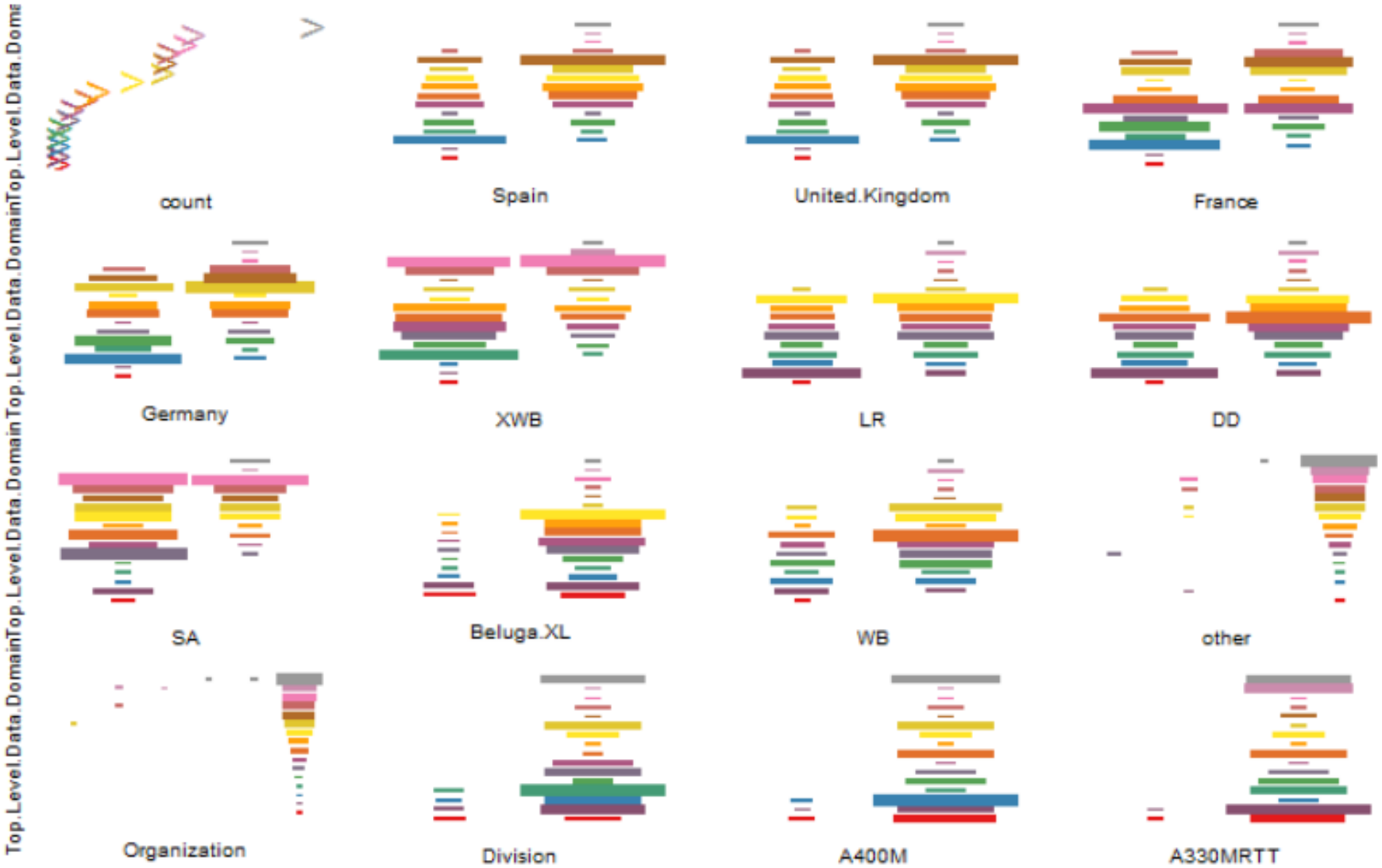
```{r data-frame summary}
inj_df %>% dplyr::select(-matches("description.*"),-matches(".*path*"))  %>% summary
```

```
                                                                                       Demand_code   Update.Date                                     Project
  scq wt_uc2 / uc3_non conformity_sap phl_31                                                :  5   Min.   :2017-02-01 00:00:00   quantum              :108    24
  quantum_kpi industrialization and root cause analysis_work order type (field name autr)_sap_42:  4   1st Qu.:2017-02-28 00:00:00   pc/spc               : 89    43
  fusion_sa e2e dashboard_techrequest dossier _techrequest_15                               :  3   Median :2017-04-12 00:00:00   post (supply chain): 40    44
  pc/spc_saint eloi_control plan_hive_24                                                    :  3   Mean   :2017-04-15 23:54:13   explorer a350        : 38    15
  pc/spc_saint eloi_measurement results_hive_24                                             :  3   3rd Qu.:2017-04-24 00:00:00   fusion               : 37    37
  (Other)                                                                                   :160   Max.   :2017-10-24 00:00:00   (Other)              :187   (Othe
  NA's                                                                                      :330   NA's   :9                     NA's                 :  9   NA's
                                         Use.case                                                                       Requestor.name   Priority
  sa e2e dashboard                           : 64   hugo c[redacted]                                                          :107    0     :  3
  uc2 / uc3                                  : 15   phillipe m[redacted] / emma l[redacted] / pierre [redacted] / nadeige [redacted]      : 79    1     : 49
  saint eloi                                 : 10   deniz [redacted] / amay s[redacted]                                     : 49    1 to n:  1
  kpi industrialization and root cause analysis:  9   amay s[redacted] / jeremy [redacted]                                 : 38    2     : 90
  ordering parameters optimization           :  8   mathieu z[redacted]/deniz [redacted] / amay s[redacted]: 30    3     :  1
  (Other)                                    : 19   (Other)                                                                   :121    NA's  :364
  NA's                                       :383   NA's                                                                      : 84
                            Source.Data.Base                              Authorative.data.source   Top.Level.Data.Domain                                       Data.Domain
  sap phl                              :166    airsupply                              : 28    manufacturing:125    industrial execution             : 81    machi
  core all                             : 28    acpng                                  : 12    transversal  : 67    material                         : 62    mater
  isaim                                : 26    sap pgi / sap pda / sap apd / sap spa:  7    procurement  : 55    quality                          : 49    work
  dms                                  :  9    pea                                    :  6    engineering  : 53    production control & scheduling: 31    purch
  sap pgi / sap pda / sap apd / sap spa:  7    (bw rejection rate)                    :  3    quality      : 52    ordering                         : 30    non c
  (Other)                              :219    (Other)                                : 19    (Other)      : 93    (Other)                          :196    (Othe
  NA's                                 : 53    NA's                                   :433    NA's         : 63    NA's                             : 59    NA's
              Sub.Concept                    Requested.dataset                                                                      Business.object.attribute
  attached documents  :  5   data captured by machine: 57   aedat / sydat / bedat\r\nwaers\r\nkdatb\r\nkdate\r\nekorg\r\nebeln\r\n                          :
  supplier performance:  5   work order              : 14   bima reference for selected msn (g9 savings)                                                 :
  missing part        :  4   measurement results     : 13   ebeln\r\nwerks\r\npeinh\r\nbprme\r\nplifz\r\nmenge\r\nebelp\r\npeinh\r\nnetwr\r\nmeins\r\nbstyp\r\n:
  drawing sheet       :  3   non conformity          :  9   plifz\r\nbstmi\r\neisbe\r\nminbe\r\nshzet\r\n\r\n                                              :
  post flight report  :  3   concession              :  5   production date (per plant, per msn)                                                         :
  (Other)             : 74   (Other)                 :373   (Other)                                                                                      :
  NA's                :414   NA's                    : 37   NA's                                                                                         :50
           Division         WB         SA          LR          XWB         DD          Beluga.XL    A400M      A330MRTT                       other           France     Germa
  airbus commercial: 16   x    :110   x    :371   x    :213   x    :336   x    :235   x    : 41   x    :  6   x    :  2   all pn except military only:  3   x    :296   x    :
  NA's             :492   NA's:398   NA's:137   NA's:295   NA's:172   NA's:273   NA's:467   NA's:502   NA's:506   involved suppliers         : 12   NA's:212   NA's:
                                                                                                                 x                          :  1
                                                                                                                 NA's                       :492
```

Un grand Dataset ? (2) R plotluck()

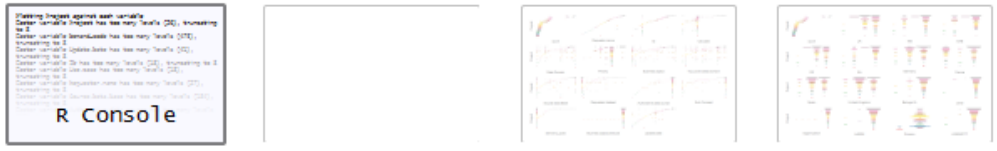


Un grand Dataset ? (2) R plotluck()



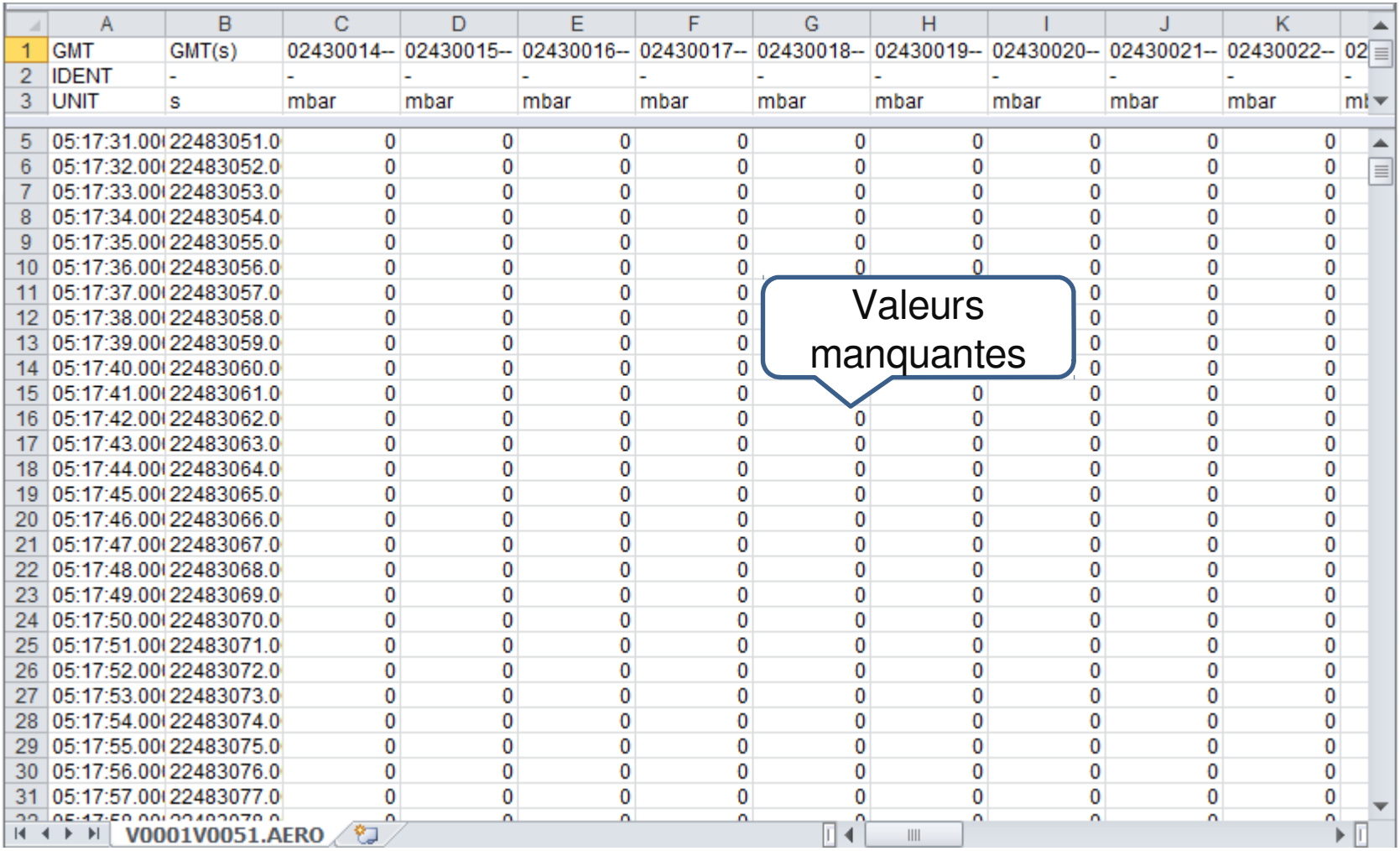
Un grand Dataset ? (2) R plotluck()

```
# Analyse data sizing and quality (missing values, skewness, structure, summaries, )
```{r single-file exploratory graph}
inj_df %>% select(1:15) %>% plotluck(formula=Project~.,opts= plotluck.options(verbose=TRUE))
inj_df %>% select(3,16:30) %>% plotluck(formula=Project~.,opts= plotluck.options(verbose=TRUE))
```
```




```
Plotting Project against each variable
Factor variable Project has too many levels (26), truncating to 8
Factor variable Demand_code has too many levels (473), truncating to 8
Factor variable Update.Date has too many levels (41), truncating to 8
Factor variable ID has too many levels (18), truncating to 8
Factor variable Use.case has too many levels (18), truncating to 8
Factor variable Requestor.name has too many levels (27), truncating to 8
Factor variable Source.Data.Base has too many levels (184), truncating to 8
Factor variable Authorative.data.source has too many levels (19), truncating to 8
Factor variable Top.Level.Data.Domain has too many levels (18), truncating to 8
Factor variable Data.Domain has too many levels (54), truncating to 8
Factor variable Business.object has too many levels (120), truncating to 8
Factor variable Sub.Concept has too many levels (74), truncating to 8
Factor variable Requested.dataset has too many levels (339), truncating to 8
Factor variable Business.object.attribute has too many levels (10), truncating to
8
Ordering variables according to conditional entropy:
                        var  cond.ent
                    Project 0.0000000
              Requestor.name 0.2620120
                         ID 0.5851395
                   Use.case 1.7324928
                Data.Domain 1.7957749
                   Priority 1.8850834
             Business.object 1.9762719
      Top.Level.Data.Domain 2.0087406
           Source.Data.Base 2.2269120
```

Un grand Dataset ? (3) csv !



	A	B	C	D	E	F	G	H	I	J	K	
1	GMT	GMT(s)	02430014--	02430015--	02430016--	02430017--	02430018--	02430019--	02430020--	02430021--	02430022--	02
2	IDENT	-	-	-	-	-	-	-	-	-	-	-
3	UNIT	s	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mb
5	05:17:31.00	22483051.0	0	0	0	0	0	0	0	0	0	
6	05:17:32.00	22483052.0	0	0	0	0	0	0	0	0	0	
7	05:17:33.00	22483053.0	0	0	0	0	0	0	0	0	0	
8	05:17:34.00	22483054.0	0	0	0	0	0	0	0	0	0	
9	05:17:35.00	22483055.0	0	0	0	0	0	0	0	0	0	
10	05:17:36.00	22483056.0	0	0	0	0	0	0	0	0	0	
11	05:17:37.00	22483057.0	0	0	0	0			0	0	0	
12	05:17:38.00	22483058.0	0	0	0	0			0	0	0	
13	05:17:39.00	22483059.0	0	0	0	0			0	0	0	
14	05:17:40.00	22483060.0	0	0	0	0			0	0	0	
15	05:17:41.00	22483061.0	0	0	0	0		0	0	0	0	
16	05:17:42.00	22483062.0	0	0	0	0	0	0	0	0	0	
17	05:17:43.00	22483063.0	0	0	0	0	0	0	0	0	0	
18	05:17:44.00	22483064.0	0	0	0	0	0	0	0	0	0	
19	05:17:45.00	22483065.0	0	0	0	0	0	0	0	0	0	
20	05:17:46.00	22483066.0	0	0	0	0	0	0	0	0	0	
21	05:17:47.00	22483067.0	0	0	0	0	0	0	0	0	0	
22	05:17:48.00	22483068.0	0	0	0	0	0	0	0	0	0	
23	05:17:49.00	22483069.0	0	0	0	0	0	0	0	0	0	
24	05:17:50.00	22483070.0	0	0	0	0	0	0	0	0	0	
25	05:17:51.00	22483071.0	0	0	0	0	0	0	0	0	0	
26	05:17:52.00	22483072.0	0	0	0	0	0	0	0	0	0	
27	05:17:53.00	22483073.0	0	0	0	0	0	0	0	0	0	
28	05:17:54.00	22483074.0	0	0	0	0	0	0	0	0	0	
29	05:17:55.00	22483075.0	0	0	0	0	0	0	0	0	0	
30	05:17:56.00	22483076.0	0	0	0	0	0	0	0	0	0	
31	05:17:57.00	22483077.0	0	0	0	0	0	0	0	0	0	

Valeurs manquantes

V0001V0051.AERO

Un grand Dataset ? (3) csv !

	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV
1	02440071--	02440072--	02440073--	02440074--	02470050--	02470051--	02470052--	02470053--	02470054--	02470055--	02470056--	02470057--	02220001--	89000002--	-11033301	74310169X4	02U01020--	88U02310--	27P100098/	27U-M135--	02U01152--	02U01151--
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	NZEV	N1_IND_A1	ALPHAPQCI	-	SLATPOS_8	SFCONF	VCCUB	ZPCUB
3	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	mbar	-	G	%	deg A	kg	deg A	-	kt	ft
2533	-0.0854492	-0.1068115	-0.1495361	-0.1495361	-0.1068115	-0.1068115	-0.1068115	-0.0854492	24.737549	-0.1708984	-0.1708984	-0.1495361	996.77496	0	1.0025001	19.41	13.642232	174866.58	18.123047	2.991559	0.06561187	454.3425
2534	-0.0854492	-0.0854492	-0.1495361	-0.1495361	-0.1068115	-0.1068115	-0.0854492	-0.0854492	24.737549	-0.1495361	-0.1495361	-0.1281738	996.77496	0	1.0025001	19.41	13.642232	174866.09	18.123047	2.991559	0.06561187	454.3425
2535	-0.0854492	-0.1281738	-0.1708984	-0.1708984	-0.1068115	-0.0640869	-0.0427246	-0.0427246	24.758911	-0.1708984	-0.1495361	-0.1281738	996.74298	0	1.00625	19.51	13.642232	174865.58	18.123047	2.991559	0.0656	454.3425
2536	-0.1068115	-0.1495361	-0.1922607	-0.1708984	-0.1068115	-0.1068115	-0.0640869	-0.0640869	24.737549	-0.1495361	-0.1495361	-0.1495361	996.74298	0	1.0025001	19.51	13.642232	174865.0				336
2537	-0.1068115	-0.1068115	-0.1495361	-0.1495361	-0.0854492	-0.0854492	-0.0854492	-0.0640869	24.737549	-0.1281738	-0.1495361	-0.1068115	996.74298	0	1.00875	19.51	13.646704	174864.				36
2538	-0.0854492	-0.1068115	-0.1708984	-0.1495361	-0.0854492	-0.1068115	-0.1068115	-0.1068115	24.737549	-0.1281738	-0.1495361	-0.1495361	996.74298	0	1.00125	19.51	13.646704	174864.				4517
2539	-0.1068115	-0.1068115	-0.1495361	-0.1281738	-0.1281738	-0.1068115	-0.0640869	-0.0427246	24.737549	-0.1281738	-0.1495361	-0.1281738	996.74298	0	1.0025001	19.51	13.646704	174863.				7493
2540	-0.1495361	-0.1281738	-0.1922607	-0.1708984	-0.0854492	-0.0854492	-0.0640869	-0.0640869	24.716187	-0.1495361	-0.1495361	-0.1495361	996.74298	0	1.00625	19.51	13.646704	174862.				425
2541	-0.1495361	-0.1281738	-0.1495361	-0.1495361	-0.0640869	-0.0640869	-0.0640869	-0.0854492	24.716187	-0.1495361	-0.1495361	-0.1495361	996.74298	0	1.01125	19.41	13.646704	174862.				7493
2542	-0.1281738	-0.1281738	-0.1495361	-0.1495361	-0.0427246	-0.0213623	-0.0213623	-0.0427246	24.737549	-0.1495361	-0.1495361	-0.1281738	996.74298	0	1.005	19.41	13.646704	174861.				7493
2543	-0.1068115	-0.1068115	-0.1068115	-0.1068115	0	0	0	-0.0213623	24.758911	-0.1068115	-0.1281738	-0.1068115	996.74298	0	0.995	19.41	13.646704	174861.36	18.123047	2.991559	0.06561239	453.91003
2544	-0.0640869	-0.0640869	-0.1281738	-0.1068115	-0.0640869	-0.0640869	-0.0640869	-0.0427246	24.737549	-0.1068115	-0.1068115	-0.1068115	996.74298	0	1.0025001	19.41	13.646704	174860.88	18.123047	2.991559	0.06561239	453.91003
2545	-0.0854492	-0.0854492	-0.1495361	-0.1281738	-0.1068115	-0.1281738	-0.1281738	-0.1281738	24.694824	-0.1495361	-0.1281738	-0.1281738	996.74298	0	0.995	19.41	13.646704	174860.36	18.123047	2.991559	0.06561187	454.3425
2546	-0.0854492	-0.1068115	-0.1281738	-0.1495361	-0.0854492	-0.0854492	-0.0854492	-0.0854492	24.716187	-0.1495361	-0.1281738	-0.1068115	996.74298	0	0.99874997	19.41	13.66585	174859.89	18.123047	2.991559	0.06561187	454.3425
2547	-0.1068115	-0.1068115	-0.1708984	-0.1922607	-0.0640869	-0.0640869	-0.0640869	-0.0854492	24.737549	-0.1922607	-0.2349853	-0.2136230	996.74298	0	1.00625	19.41	13.66585	174859.38	18.123047	2.991559	0.06561136	454.77493
2548	-0.1495361	-0.1495361	-0.1922607	-0.1708984	-0.0213623	-0.0427246	-0.0213623	-0.0640869	24.758911	-0.1708984	-0.1922607	-0.1708984	996.74298	0	1.0025001	19.41	13.66585	174858.89	18.123047	2.991559	0.06561136	454.77493
2549	-0.1068115	-0.1281738	-0.1495361	-0.1495361	-0.0427246	-0.0427246	-0.0213623	-0.0427246	24.758911	-0.1281738	-0.1281738	-0.1068115	996.74298	0	1.00625	19.41	13.66585	174858.39	18.123047	2.991559	0.06561290	453.4776
2550	-0.1281738	-0.1068115	-0.1708984	-0.1281738	-0.0427246	-0.0427246	-0.0213623	-0.0427246	24.758911	-0.1495361	-0.1495361	-0.1495361	996.74298	0	1.00625	19.41	13.66585	174857.89	18.123047	2.991559	0.06561239	453.91003
2551	-0.1068115	-0.1068115	-0.1495361	-0.1281738	-0.0427246	-0.0427246	-0.0213623	-0.0427246	24.758911	-0.1281738	-0.1281738	-0.1281738	996.74298	0	1.00625	19.41	13.66585	174857.14	18.123047	2.991559	0.06561239	453.91003
2552	-0.0640869	-0.0640869	-0.1281738	-0.1068115	-0.0640869	-0.0640869	-0.0640869	-0.0427246	24.780273	-0.0854492	-0.0854492	-0.0854492	996.77496	0	0.9975	19.51	13.66585	174856.66	18.123047	2.991559	0.06561187	454.3425
2553	-0.0640869	-0.0854492	-0.1281738	-0.1281738	-0.0640869	-0.0640869	-0.0640869	-0.0640869	24.758911	-0.1281738	-0.1495361	-0.1281738	996.77496	0	1.00375	19.610001	13.66585	174856.17	18.123047	2.991559	0.06561239	453.91003
2554	-0.1068115	-0.1281738	-0.1708984	-0.1495361	-0.0640869	-0.0640869	-0.0427246	-0.0427246	24.737549	-0.1495361	-0.1708984	-0.1495361	996.74298	0	0.99874997	19.710001	13.670331	174855.67	18.123047	2.991559	0.06561239	453.91003
2555	-0.1068115	-0.1068115	-0.1495361	-0.1495361	-0.0640869	-0.0640869	-0.0427246	-0.0640869	24.737549	-0.1495361	-0.1495361	-0.1281738	996.77496	0	1.01375	19.710001	13.670331	174855.17	18.123047	2.991559	0.06561239	453.91003
2556	-0.1068115	-0.1068115	-0.1495361	-0.1281738	-0.0854492	-0.0854492	-0.0640869	-0.0640869	24.737549	-0.1281738	-0.1281738	-0.1281738	996.74298	0	0.99624997	19.710001	13.670331	174854.69	18.123047	2.991559	0.06561239	453.91003
2557	-0.0854492	-0.0854492	-0.1495361	-0.1495361	-0.0427246	-0.0427246	-0.0427246	-0.0640869	24.758911	-0.1708984	-0.1708984	-0.1495361	996.77496	0	0.99624997	19.710001	13.670331	174854.2	18.123047	2.991559	0.06561239	453.91003
2558	-0.1068115	-0.1281738	-0.1708984	-0.1708984	-0.0213623	-0.0213623	0	-0.0427246	24.758911	-0.1708984	-0.1495361	-0.1068115	996.74298	0	1	19.710001	13.651175	174853.72	18.123047	2.991559	0.06561239	453.91003
2559	-0.1495361	-0.1281738	-0.1495361	-0.1281738	-0.0213623	-0.0213623	0	-0.0427246	24.758911	-0.1281738	-0.1281738	-0.1068115	996.74298	0	1.005	19.610001	13.651175	174853.22	18.123047	2.991559	0.06561239	453.91003
2560	-0.1068115	-0.1068115	-0.1068115	-0.0854492	-0.0213623	-0.0213623	0	0	24.780273	-0.0854492	-0.0854492	-0.0640869	996.71094	0	1.005	19.610001	13.651546	174852.73	18.123047	2.991559	3.3491125	454.74036
2561	-0.1068115	-0.0854492	-0.1068115	-0.1068115	0	-0.0213623	0	-0.0213623	24.801636	-0.0854492	-0.0854492	-0.0854492	996.71094	0	0.9975	19.610001	13.651546	174852.22	18.123047	2.991559	3.3491125	454.74036
2562	-0.0854492	-0.1068115	-0.1495361	-0.1281738	-0.0213623	-0.0213623	-0.0213623	-0.0213623	24.780273	-0.1281738	-0.1281738	-0.1068115	996.74298	0	1.0025001	19.51	13.651175	174851.75	18.123047	2.991559	0.06561136	454.77493
2563	-0.1281738	-0.1495361	-0.1922607	-0.1708984	-0.0213623	-0.0213623	-0.0213623	-0.0427246	24.780273	-0.1495361	-0.1495361	-0.1495361	996.74298	0	1.0025001	19.41	13.651546	174851.02	18.123047	2.991559	3.3491125	454.74036
2564	-0.1068115	-0.0854492	-0.1281738	-0.1068115	-0.0213623	0	0	-0.0213623	24.801636	-0.1068115	-0.0854492	-0.0854492	996.71094	0	1.00125	19.41	13.651336	174850.52	18.123047	2.991559	1.4830265	454.336
2565	0	-0.0427246	-0.1281738	-0.0854492	-0.0213623	-0.0213623	0	-0.0213623	24.801636	-0.0854492	-0.0854492	-0.0640869	996.71094	0	0.9975	19.41	13.651336	174850.03	18.123047	2.991559	1.4830265	454.336
2566	-0.0427246	-0.0427246	-0.1708984	-0.1495361	-0.0213623	-0.0213623	0	-0.0213623	24.801636	-0.1281738	-0.1068115	-0.1068115	996.71094	0	1.01125	19.41	13.651175	174849.55	18.123047	2.991559	0.06561239	453.90356
2567	-0.0640869	-0.0854492	-0.1281738	-0.1281738	-0.0213623	-0.0213623	0	0	24.801636	-0.1281738	-0.1281738	-0.1068115	996.71094	0	1.01125	19.310001	13.651678	174849.05	18.123047	2.991559	4.5062943	455.14685
2568	-0.0854492	-0.1281738	-0.1495361	-0.1495361	0	0	0	0	24.822998	-0.1281738	-0.1281738	-0.1068115	996.71094	0	1.00125	19.310001	13.651546	174848.56	18.123047	2.991559	3.3490598	455.60526
2569	-0.1068115	-0.1281738	-0.1281738	-0.0854492	0.02136230	0.02136230	0.04272460	0.02136230	24.84436	-0.1068115	-0.1068115	-0.0854492	996.67896	0	1.0025001	19.310001	13.651678	174848.06	18.123047	2.991559	4.5062943	455.14685
2570	-0.2136230	-0.1708984	-0.1495361	-0.1068115	0	0.02136230	0.02136230	0	24.822998	-0.1068115	-0.1068115	-0.0854492	996.71094	0	1.00375	19.210001	13.651678	174847.56	18.123047	2.991559	4.5062943	455.14685
2571	-0.1495361	-0.1922607	-0.1495361	-0.1281738	0	0	0.02136230	0	24.822998	-0.1281738	-0.1068115	-0.0854492	996.71094	0	1.0025001	19.210001	13.651336	174847.08	18.123047	2.991559	1.4830031	455.20093
2572	-0.1068115	-0.1068115	-0.1495361	-0.1281738	0	0	0	0	24.822998	-0.1068115	-0.1068115	-0.0854492	996.71094	0	1.00375	19.310001	13.651175	174846.62	18.123047	2.991559	0.06561136	454.77493
2573	-0.1495361	-0.1708984	-0.1495361	-0.1281738	-0.0213623	-0.0213623	-0.0213623	-0.0213623	24.822998	-0.1281738	-0.1281738	-0.1068115	996.71094	0	1.005	19.310001	13.651175	174846.11	18.123047	2.991559	0.06561187	454.3425
2574	-0.1068115	-0.1281738	-0.1922607	-0.1708984	-0.0213623	-0.0213623	-0.0213623	-0.0427246	24.822998	-0.1495361	-0.1495361	-0.1281738	996.71094	0	1.0025001	19.41	13.651175	174845.61	18.123047	2.991559	0.06561239	453.91003
2575	-0.1708984	-0.1922607	-0.1708984	-0.1495361	-0.0213623	-0.0213623	-0.0213623	-0.0427246	24.822998	-0.1708984	-0.1922607	-0.1495361	996.71094	0	1.0025001	19.41	13.651175	174844.88	18.123047	2.991559	0.06561290	453.4776
2576	-0.1281738	-0.1281738	-0.1495361	-0.1281738	-0.0427246	-0.0427246	-0.0427246	-0.0427246	24.801636	-0.1495361	-0.1281738	-0.1068115	996.71094	0	1.00125	19.41	13.651175	174844.39	18.123047	2.991559	0.06561239	453.91003
2577	-0.1281738	-0.1281738	-0.1708984	-0.1495361	-0.0640869	-0.0427246	-0.0427246	-0.0427246	24.822998	-0.1281738	-0.1281738	-0.1281738	996.74298	0	1.00375	19.41	13.651175	174843.91	18.123047	2.991559	0.06561136	454.77493
2578	-0.2136230	-0.2563476	-0.1922607	-0.1708984	-0.0213623	-0.0427246	-0.0213623	-0.0427246	24.822998	-0.1708984	-0.1495361	-0.1495361	996.74298	0	1.00375	19.41	13.651175	174843.41	18.123047	2.991559	0.06561136	454.77493
2579	-0.2563476					46	-0.0427246	-0.0427246	-0.0640869	24.84436	-0.1708984	-0.1922607	-0.1495361	996.71094	0	1.00625	19.51	13.651175	174842.91	18.123047	2.991559	0.06561136	454.77493
2580	-0.1495361					23	-0.0427246	-0.0213623	-0.0427246	24.84436	-0.1281738	-0.1281738	-0.1068115	996.74298	0	1.005	19.51	13.651175	174842.42	18.123047	2.991559	0.06561187	454.3425

51 colonnes

18 000 lignes

Un grand Dataset ? (3) R summary()


```
# Initialisation

Here is the library loading required for analysis, and environment and system setup in order
for the analysis to be reproductible.
```{r libraries, include=FALSE}
library(tidyverse)
library(readxl)
library(stringr)
library(plotluck)
```

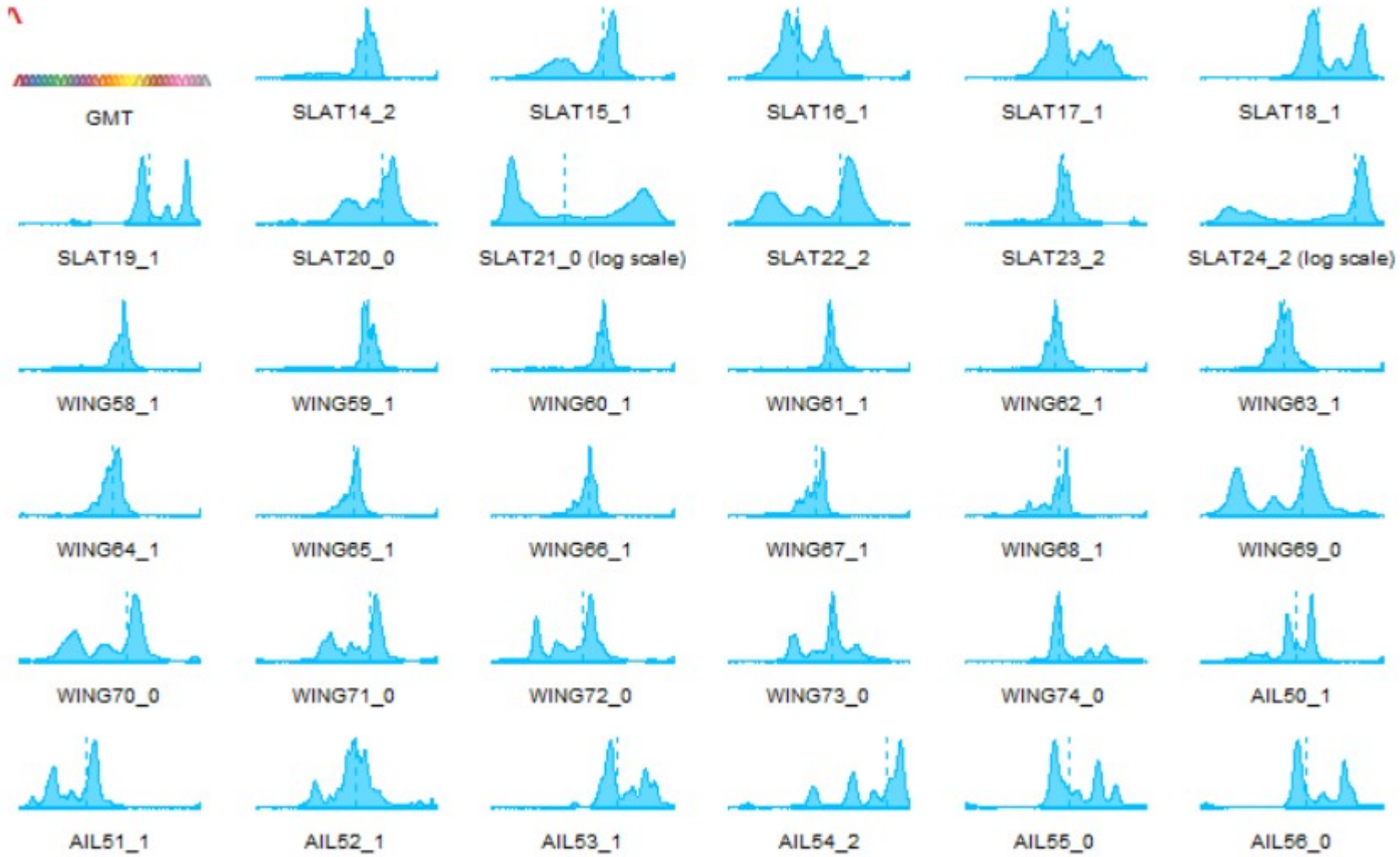
```
> load(file="scaling_aero.Rda")
> ae51<-aero[aero$flight=="V0051",]
```

```
> #plotluck is sensible to constant columns and na and doesn't understand difftime. We help the tool
with some data cleaning
> ae51<-ae51[,sapply(ae51, function(v) !all(duplicated(na.omit(v))[-1L]))] %>% mutate_at("rel_time",a
s.numeric)
```
```

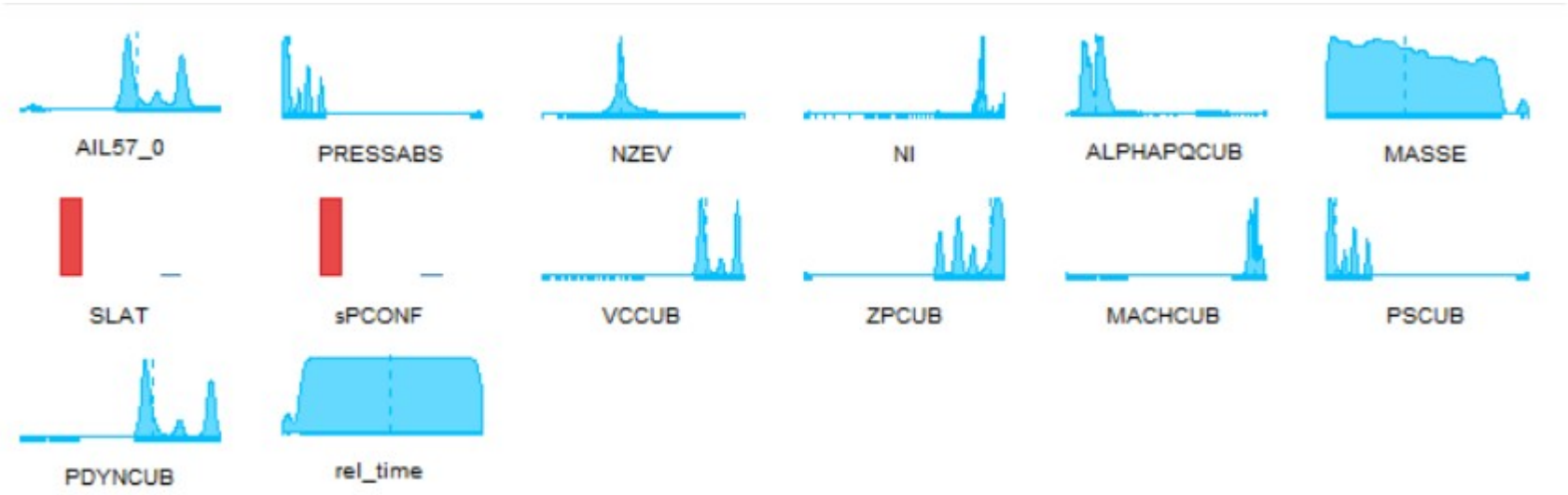
Un grand Dataset ? (3) R summary()

```
> summary(ae51)
     GMT              SLAT14_2          SLAT15_1          SLAT16_1          SLAT17_1
 Length:12332       Min.   :-149.0784   Min.   :-146.0286   Min.   :-148.379   Min.   :-134.98
 Class1:hms         1st Qu.: -63.4460   1st Qu.: -86.3049   1st Qu.:-100.923   1st Qu.: -56.40
 Class2:difftime    Median : -57.7393   Median : -56.1226   Median : -91.630   Median : -45.78
 Mode  :numeric     Mean   : -60.7744   Mean   : -65.6388   Mean   : -87.360   Mean   : -39.24
                    3rd Qu.: -53.1921   3rd Qu.: -49.0348   3rd Qu.: -71.710   3rd Qu.: -19.47
                    Max.   :   0.1221   Max.   :   0.6104   Max.   :   1.099   Max.   :  24.11
    SLAT18_1          SLAT19_1          SLAT20_0          SLAT21_0          SLAT22_2
 Min.   :-126.314   Min.   :-66.10   Min.   :-21.97   Min.   :-173.800   Min.   :-66.07
 1st Qu.:   9.186   1st Qu.: 83.22   1st Qu.: 39.39   1st Qu.: -58.198   1st Qu.:-42.69
 Median :  19.821   Median : 91.77   Median : 55.19   Median :  -3.510   Median :-13.76
 Mean   :  33.528   Mean   :104.69   Mean   : 50.65   Mean   : -21.941   Mean   :-23.10
 3rd Qu.:  65.766   3rd Qu.:136.57   3rd Qu.: 61.31   3rd Qu.:   6.744   3rd Qu.: -8.27
 Max.   :  98.268   Max.   :155.15   Max.   : 88.44   Max.   :  48.859   Max.   : 18.59
    SLAT23_2          SLAT24_2          WING58_1          WING59_1          WING60_1
 Min.   :-95.765   Min.   :-28.962   Min.   :-155.0263   Min.   :-152.2064   Min.   :-147.635
 1st Qu.:-42.695   1st Qu.: -2.960   1st Qu.: -72.8935   1st Qu.: -61.3312   1st Qu.: -59.136
 Median :-40.253   Median :  9.705   Median : -67.1204   Median : -58.2764   Median : -55.627
 Mean   :-41.026   Mean   :  5.660   Mean   : -69.9178   Mean   : -58.3169   Mean   : -56.227
 3rd Qu.:-37.324   3rd Qu.: 13.581   3rd Qu.: -64.1083   3rd Qu.: -53.5126   3rd Qu.: -53.214
 Max.   :  6.775   Max.   : 30.762   Max.   :   0.1282   Max.   :   0.1495   Max.   :   2.841
    WING61_1           WING62_1           WING63_1           WING64_1
 Min.   :-136.76147   Min.   :-123.60229   Min.   :-117.04407   Min.   :-116.91589
 1st Qu.: -61.33118   1st Qu.: -64.89868   1st Qu.: -68.23120   1st Qu.: -61.03210
 Median : -59.47266   Median : -61.69434   Median : -64.28986   Median : -56.50330
```

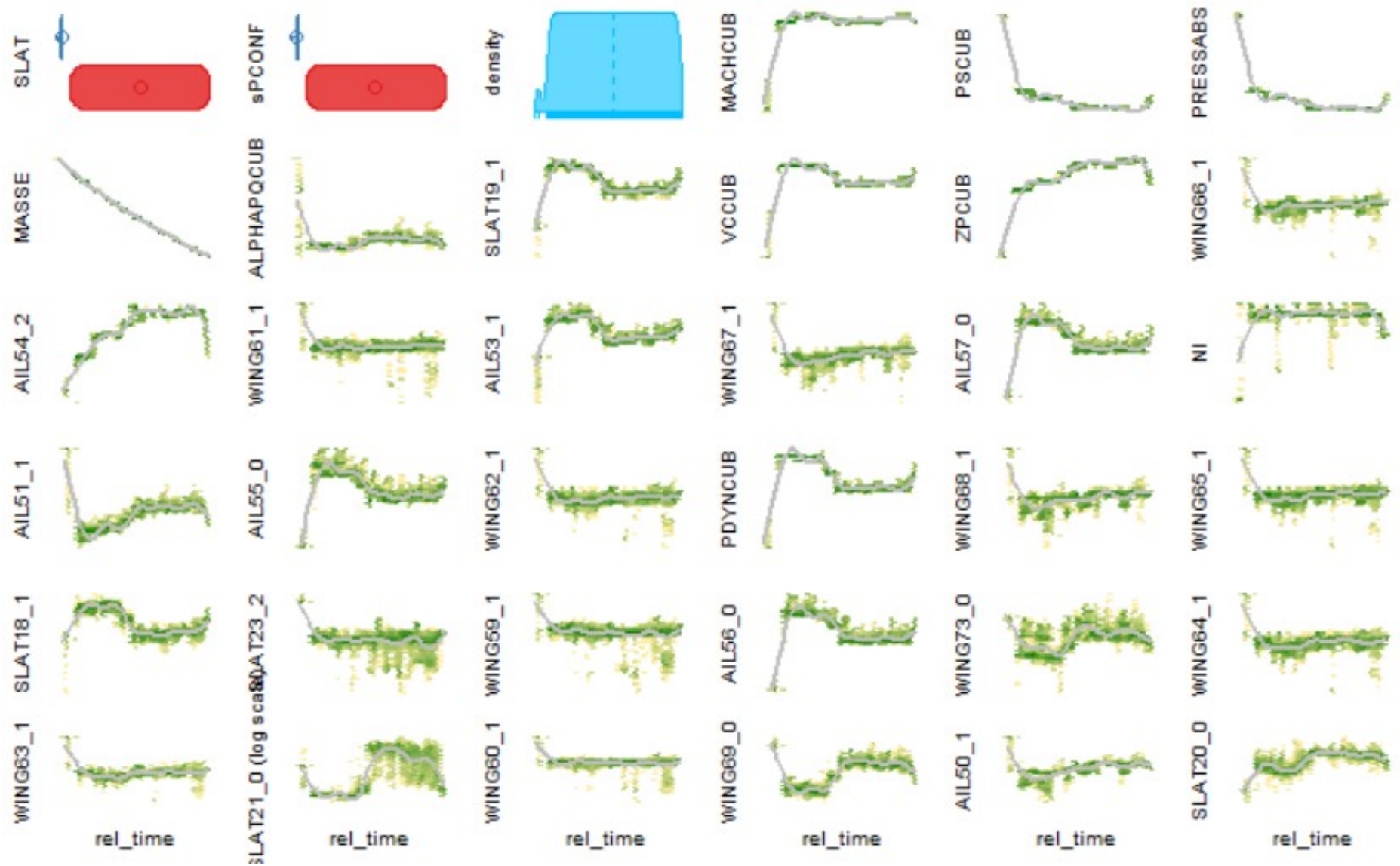
Un grand Dataset ? (3) R plotluck()



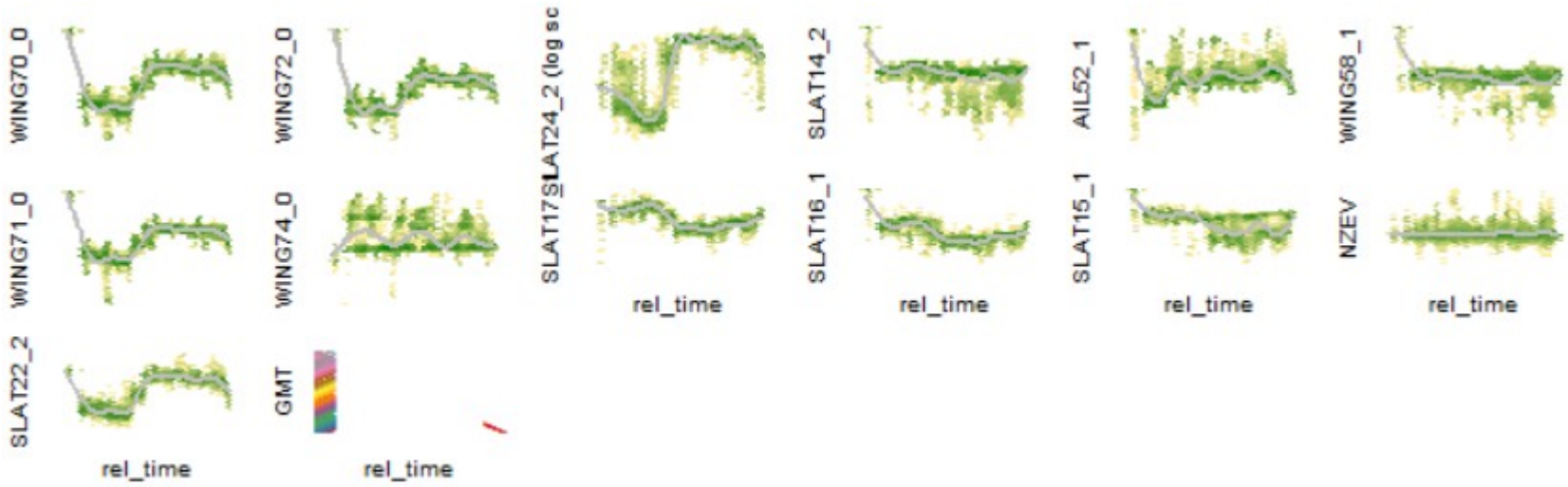
Un grand Dataset ? (3) R plotluck()



Un grand Dataset ? (3) R plotluck()



Un grand Dataset ? (3) R plotluck()



Et vous, quel est votre outil ?