





- Discussions techniques bimensuelles autour de la “mesure” et de la donnée

- Apéro Technique #01, juin 2014

Méthodes de rapatriement de données

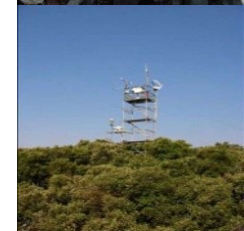
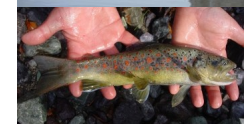
Méthodes d'alimentation de bases de données

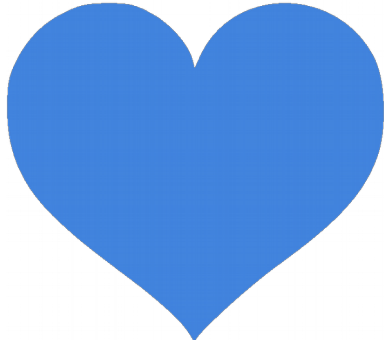
- Apéro Technique #02, oct 2014

R pour les données de la recherche #1

- Apéro Technique #03, déc 2014

Outils d'analyse en géochimie



I  R

R pour les néophytes





R

- Langage et environnement pour l'analyse, la visualisation et la modélisation de données
- Standard chez les statisticiens ... mais pas que !



<http://cran.r-project.org/>





Caractéristiques

Libre, open source, multi plateformes

Large communauté

Hyper extensible
(~ 6000 packages !)

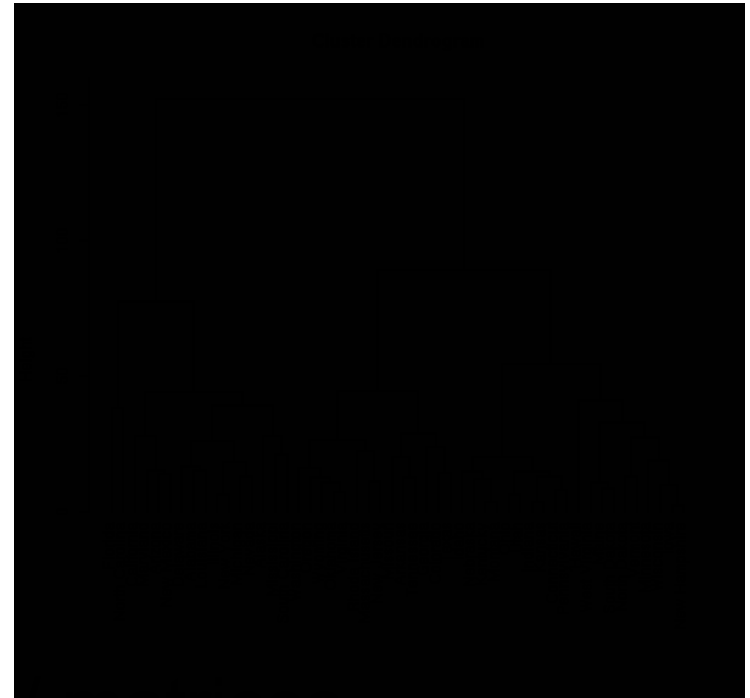
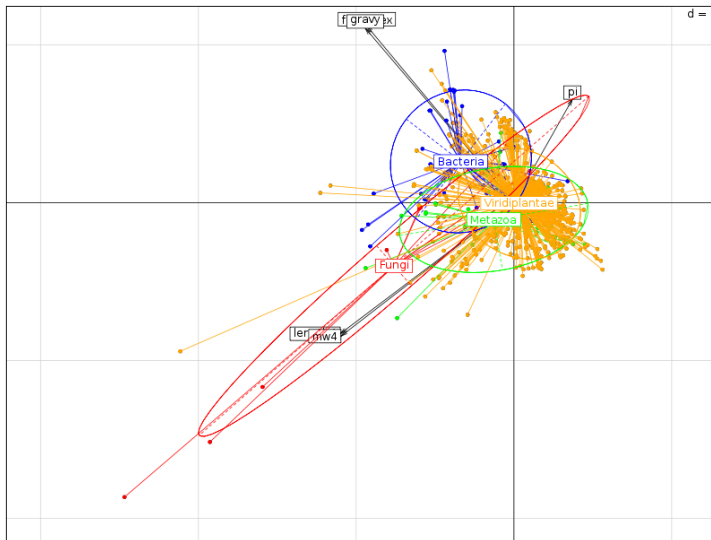
Langage interprété et interactif

Script : automatisation des traitements, reproductibilité

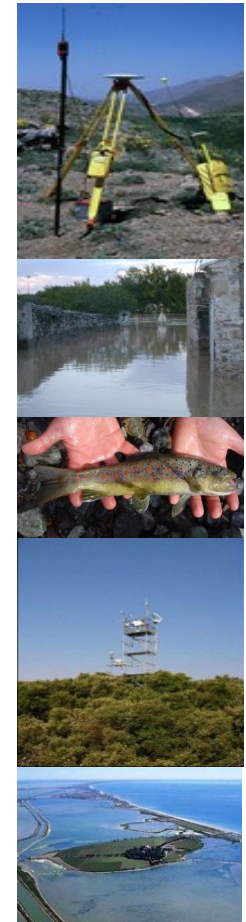


Analyses de données / calcul

- Analyses statistiques :
 - Linéaire, non linéaire, séries temporelles
 - Clustering, classification
 - Spatial, ..



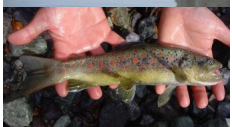
- Calcul sur les tableaux / matrices





Traitements

- Des packages toujours à la pointe
 - **Génétique** : genetics, qtl, seqinr ..
 - **Données spatiales** : plotKML, RgoogleMaps ..
 - **Web** : twittR, Shiny, Rgoogle-analytics
 - ..
- **Liste des packages par tâche** Non exhaustif !
- **Top 100 des packages** (2013)
- **Top 50 des plus utilisés** (2012)
- **Liste des packages utiles** (2010)





Graphiques

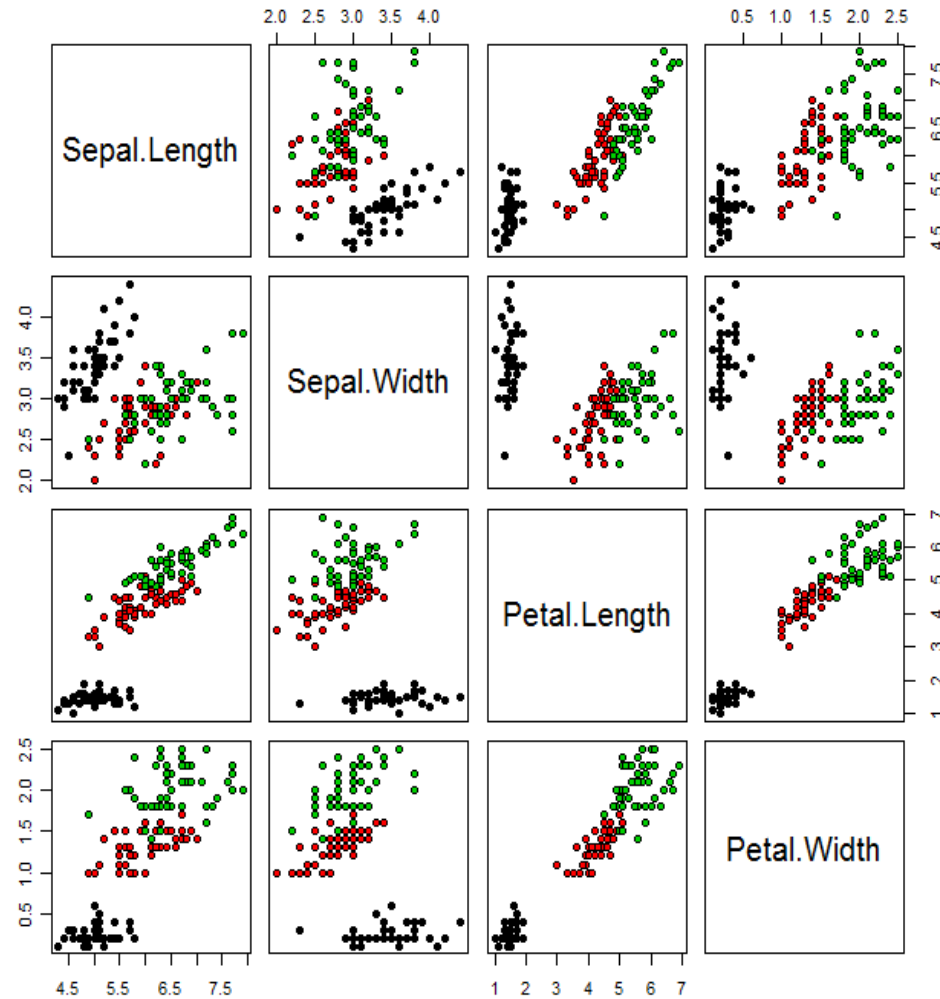
- Facilités graphiques
 - Tout est modifiable
 - Formules et symboles mathématiques
- Export : PDF, PS, JPG,





Graphiques

Nuages de point

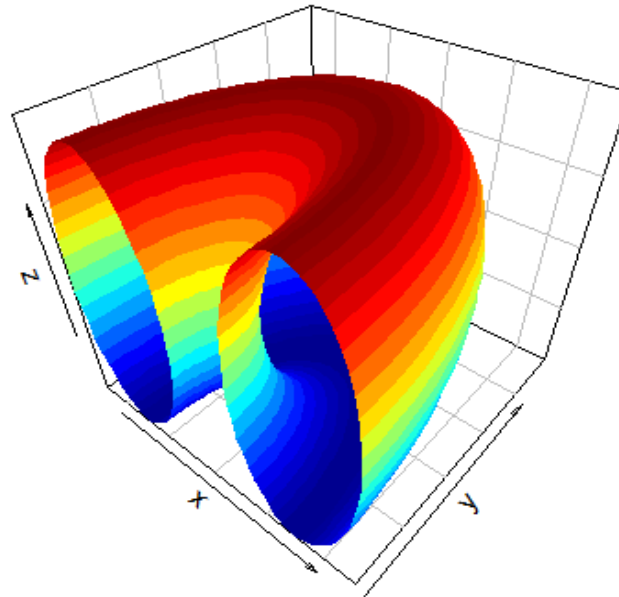




Graphiques

Perspective - 3D

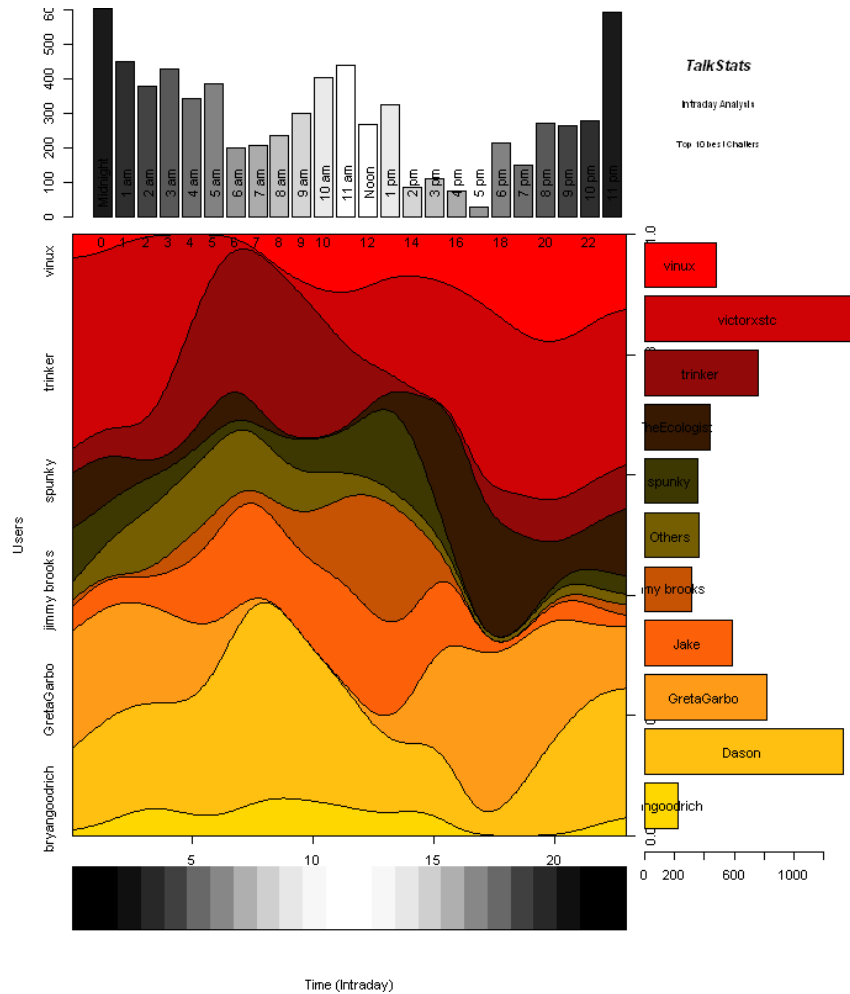
Half of a Torus





Graphiques

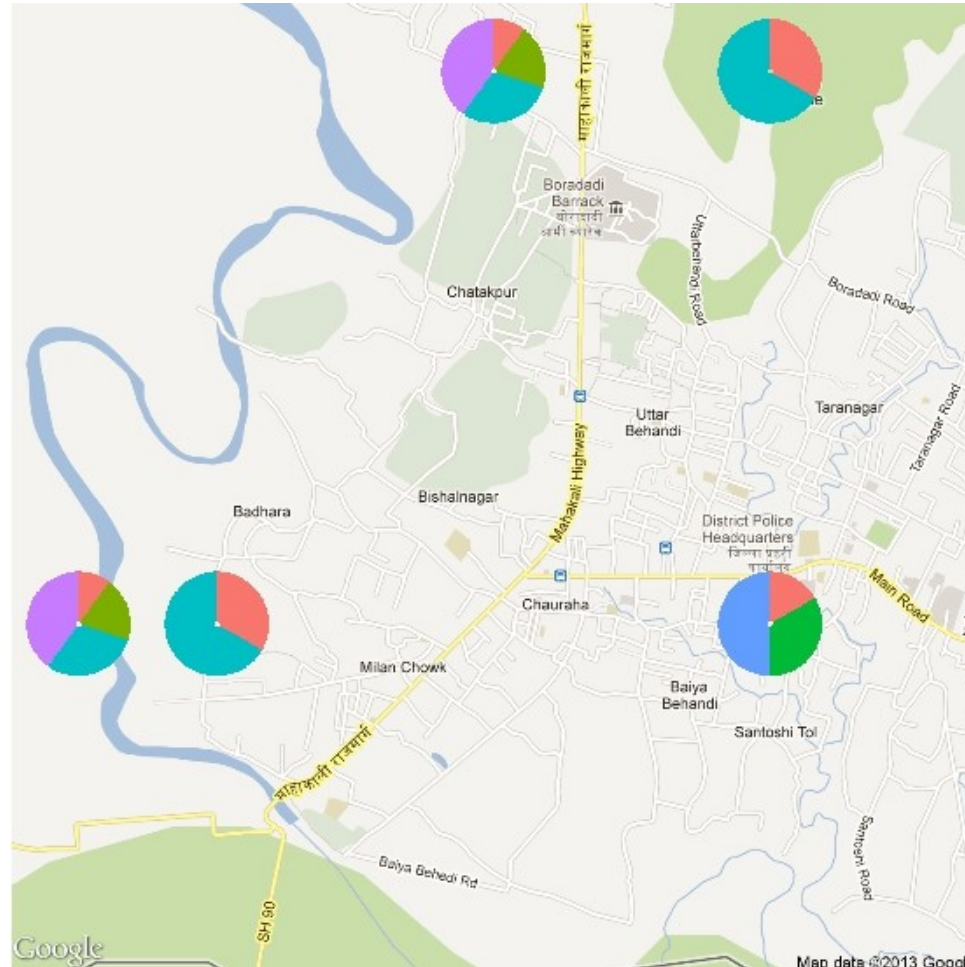
Graphs conditionnels – multi graphs





Graphiques

Cartes

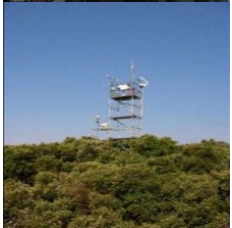
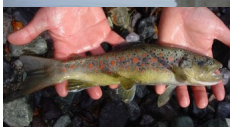


<http://rgraphgallery.blogspot.fr/search/label/map>



R et les données de la recherche

- **Import / Export**
 - Fichiers : CSV, texte, XLS, NetCDF, ..
 - Bases de données
 - Via le réseau (HTTP, FTP, web services, ...)
- Le type **dataframe** = variables x observations
- **Traitement** : Exploration, transformation, analyses, modélisation, visualisation
- **Diffusion** : Web, rapports dynamiques, ...

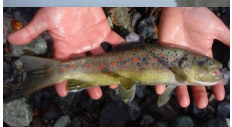




R et les données de la recherche

- Import / Export
 - Fichiers : CSV, texte, XLS, NetCDF, ..
 - Bases de données
 - Via le réseau (HTTP, FTP, web services, ...)
- Le type dataframe = variables x observations
- Traitement : Exploration, transformation, analyses, modélisation, visualisation
- Diffusion : Web, rapports dynamiques, ...

REPRODUCTIBLE





Interfaces

- RGui (Windows, Mac OS)

The screenshot displays the RGui environment with the following components:

- R Console:** A table of island names and their corresponding frequencies.
- R Help on 'hist':** Documentation for the 'hist' function, including its description and usage.
- R Graphics:** Two histograms comparing island frequencies. The left histogram is titled 'Histogram of islands' and the right is 'Histogramme des îles'. Both show a highly skewed distribution with a single dominant bar.

Island Name	Frequency
84	73
Devon	Ellesmere
21	82
Hainan	Hispaniola
13	30
Iceland	Ireland
40	33
Luzon	Madagascar
42	227
Moluccas	New Britain
29	15
New Zealand (S)	Newfoundland
58	43
Prince of Wales	Sakhalin
13	29
Spitsbergen	Sumatra
15	183
Tierra del Fuego	Timor
19	13
25	73
Europe	Greenland
3745	840
Hokkaido	Honshu
30	89
Java	Kyushu
49	14
Melville	Mindanao
16	36
New Guinea	New Zealand (N)
306	44
North America	Novaya Zemlya
9390	32
South America	Southampton
6795	16
Taiwan	Tasmania
14	26
Vancouver	Victoria
12	82

```

hist
      package:graphics      R Documentation

Histograms

Description:

The generic function 'hist' computes a histogram of the given
data values.  If 'plot=TRUE', the resulting object of class
"histogram" is plotted by 'plot.histogram', before it is
returned.

Usage:

hist(x, ...)

## Default S3 method:
hist(x, breaks = "Sturges",
     freq = NULL, probability = 'freq',
     include.lowest = TRUE, right = TRUE,
     density = NULL, angle = 45, col = NULL, border = NULL,
     main = paste("Histogram of", xname),
     xlim = range(breaks), ylim = NULL,
     xlab = xname, ylab,
     axes = TRUE, plot = TRUE, labels = FALSE,
     nclass = NULL, ...)

Arguments:

x: a vector of values for which the histogram is desired.
  
```





Interfaces

- Console

```
juliette@Port-OSU:~$ R

R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

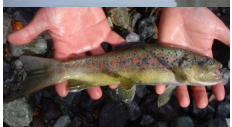
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library(XLConnect)
XLConnect 0.2-7 by Mirai Solutions GmbH
http://www.mirai-solutions.com ,
http://miraisolutions.wordpress.com
> wb <- loadWorkbook('Bureau/OREME/so_keywords.xls')
> data <- readWorksheet(wb, 1)
> head(data)
  SO.short.name      SO.long.name      Keyword.FR
1      ECOPOP      Ecologie des populations  écologie des populations
2      <NA>      <NA>      évolution
3      EVOPOP      Evolution des populations  populations
4      <NA>      <NA>      <NA>
5      EVOPOP      Evolution des populations  <NA>
6      <NA>      <NA>      <NA>

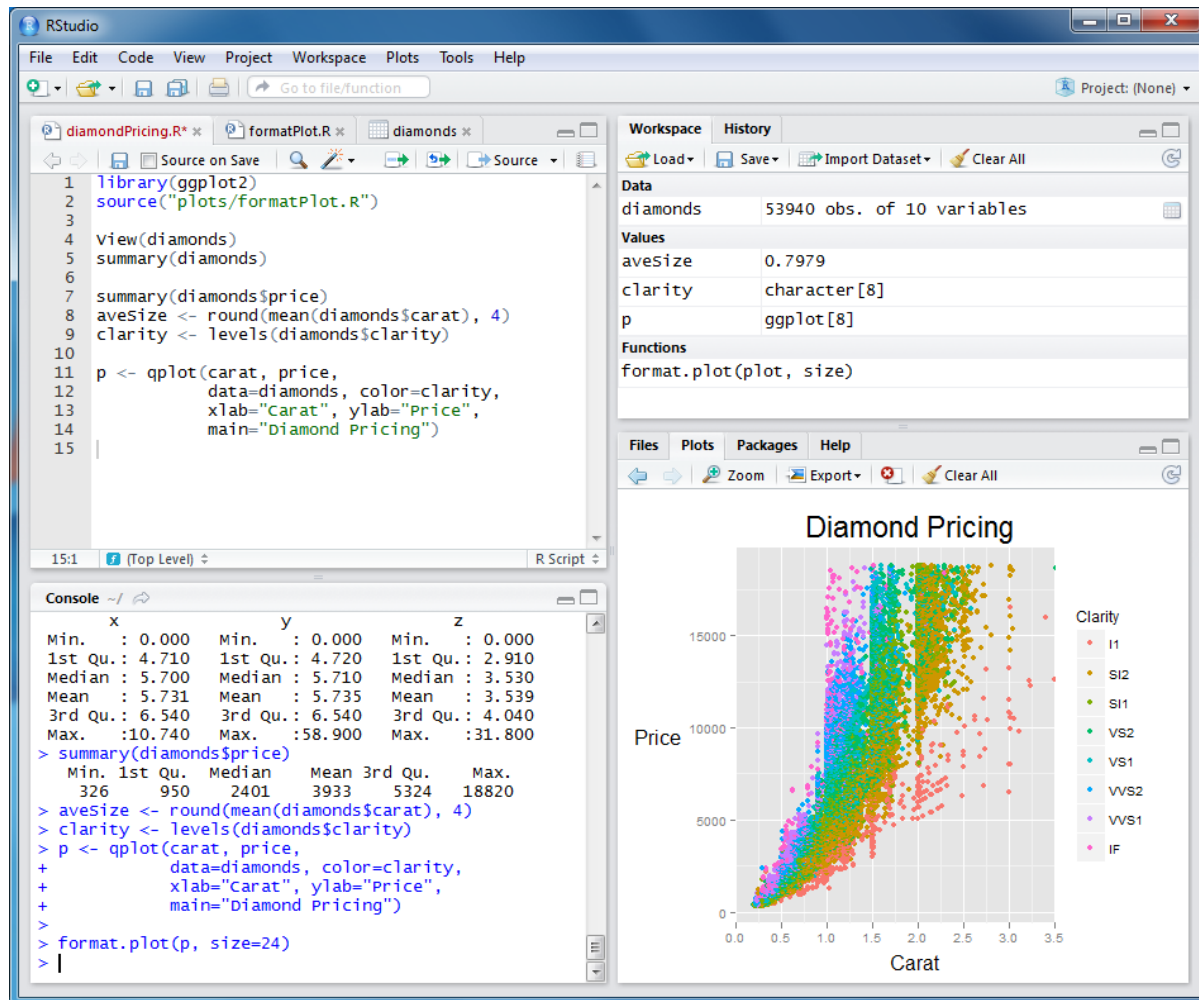
      GEMET.URI
1 http://www.eionet.europa.eu/gemet/concept/6520
2 http://www.eionet.europa.eu/gemet/concept/3026
3 http://www.eionet.europa.eu/gemet/concept/6519
4      <NA>
5      <NA>
6      <NA>

      ENVTHES.URI
1      http://vocabs.lter-europe.net/EnvThes3/82
2 http://vocabs.lter-europe.net/EnvThes3/USLterCV_186
3 http://vocabs.lter-europe.net/EnvThes3/USLterCV_439
4      <NA>
5      <NA>
6      <NA>
> |
```



RStudio

- Environnement multi plateformes



The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains an R script named `diamondPricing.R` with the following code:


```
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 View(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12            data=diamonds, color=clarity,
13            xlab="Carat", ylab="Price",
14            main="Diamond Pricing")
15
```
- Console:** Shows the execution output:


```
Min.   x: 0.000   Min.   y: 0.000   Min.   z: 0.000
1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910
Median: 5.700 Median: 5.710 Median: 3.530
Mean: 5.731 Mean: 5.735 Mean: 3.539
3rd Qu.: 6.540 3rd Qu.: 6.540 3rd Qu.: 4.040
Max.: 10.740 Max.: 58.900 Max.: 31.800
> summary(diamonds$price)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  326   950   2401   3933   5324  18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+           data=diamonds, color=clarity,
+           xlab="Carat", ylab="Price",
+           main="Diamond Pricing")
>
> format.plot(p, size=24)
>
```
- Workspace:** Shows the loaded data object `diamonds` with 53940 observations and 10 variables. The `Values` section lists `aveSize` (0.7979), `clarity` (character [8]), and `p` (ggplot [8]).
- Plots:** Displays a scatter plot titled "Diamond Pricing" showing Price (Y-axis, 0 to 15000) versus Carat (X-axis, 0.0 to 3.5). The points are colored by clarity, with a legend on the right showing categories: I1, SI2, SI1, VS2, VS1, VVS2, VVS1, and IF.

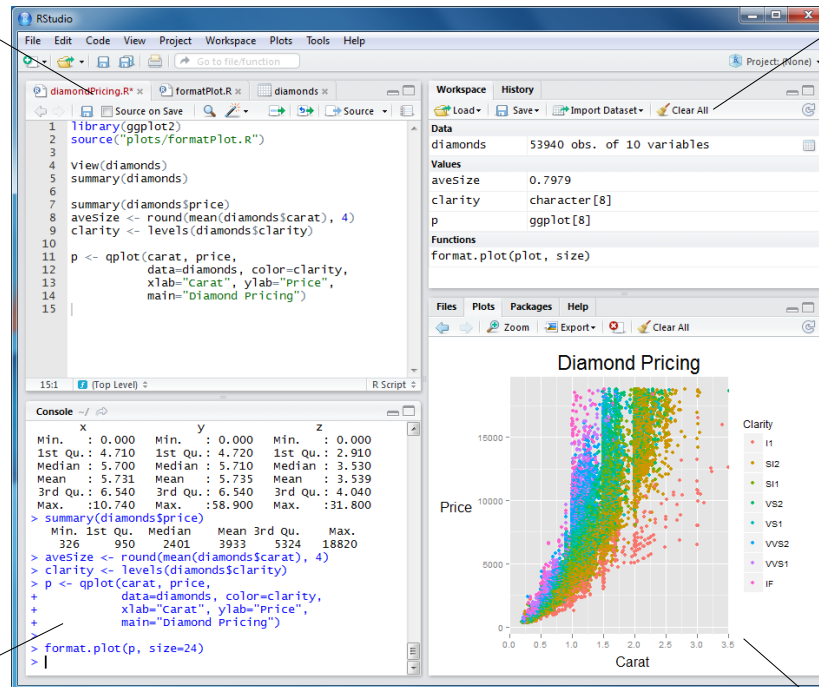




RStudio

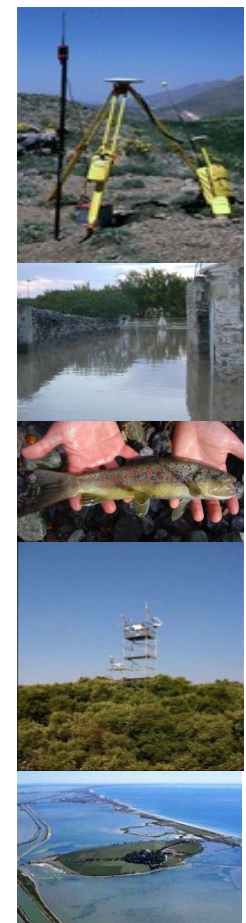
Editeur : coloration syntaxique, aide à l'exécution

Objets, Historique, **Versioning** (Git et SVN)



Console

Système de fichiers, Graphiques, Bibliothèques, Aide





RStudio

- Débuggage
- Création de “notebook” de code
- Mode serveur
- ...

medycyss_datacite_metadata_export.R

juliette

Tue Oct 7 11:55:54 2014

```
# MEDYCYSS METADATA EXPORT INTO DATACITE SCHEMA
# author : Juliette Fabre
# creation : 09/14
# last update : 09/14
```

```
# The script reads MEDYCYSS metadata tables, and for specific station and data types it formats and exports metadata into an XML file matching the Datacite format.
```

```
# Get argument
args <- commandArgs()
cleanargs <- sub("^-.*=", "", args)
#type = cleanargs[regexpr("-t=", args)>0]

# Load functions file
script_path <- dirname(sub("--file=", "", args[4]))
source(paste(script_path, "../oreme/oreme_db_v1.2_private.R", sep = ""))
```

```
# Load librairies
library(XML)
```

```
# Connect to OSU database
connexion <- connect_osu()
```

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

<http://www.rstudio.com/products/rstudio/>



Liens

- Blogs, forums, ...

- [Forum R du Cirad](#)
- [R-bloggers](#)
- [R graphs gallery](#)
- [Inside-R](#)

- Conférences :

- France : Rencontres R ([2014 Montpellier](#), 2015 Grenoble)
- Monde : *useR!* (2014 LA, [2015 Danemark](#))

