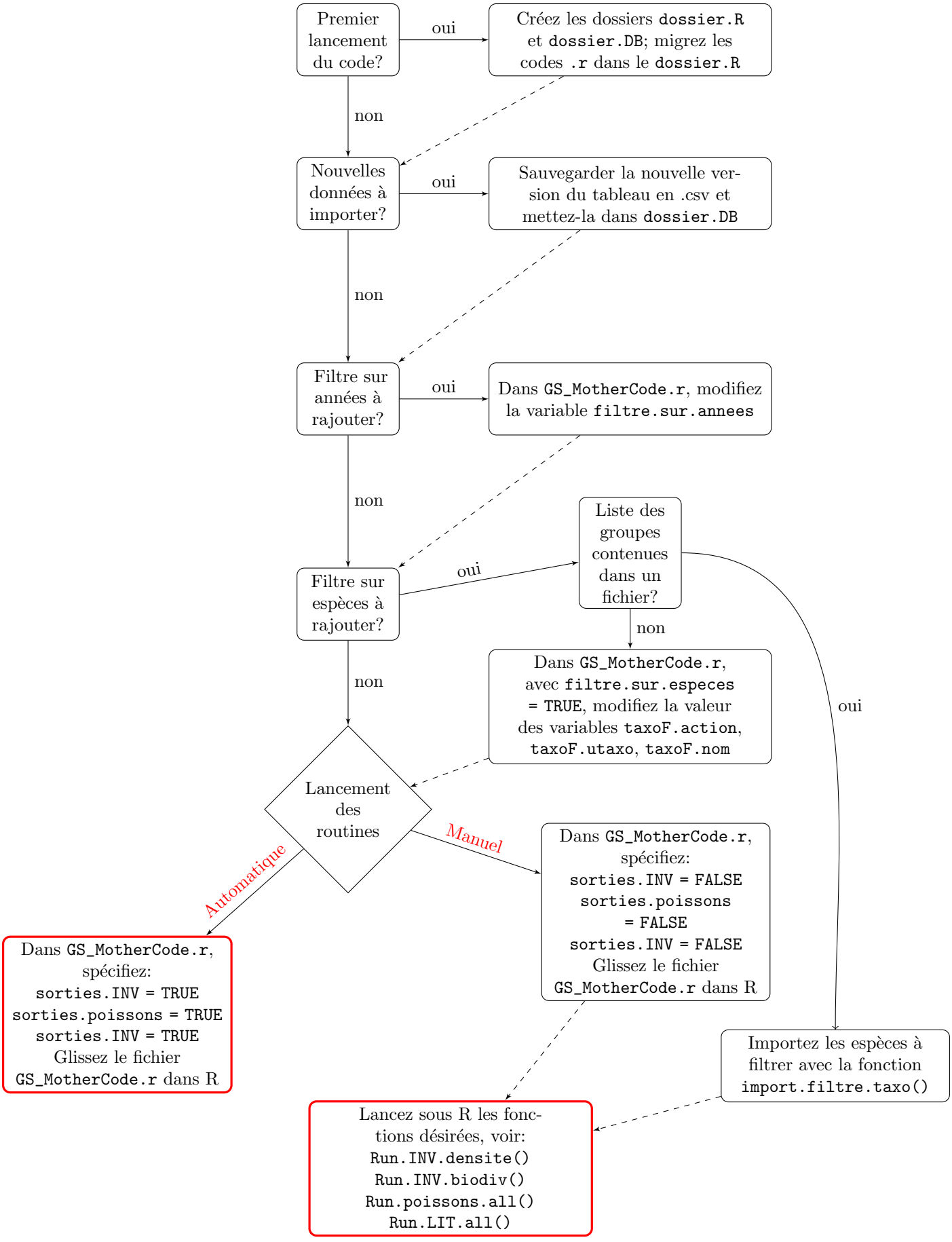
Description: Code R pour l'analyse des données d'échantillonage

Contents


Premier lancement du code?
oui
Créez les dossiers dossier.R et dossier.DB; migrez les codes .r dans le dossier.R
non
Nouvelles données à importer?
oui
Sauvegarder la nouvelle version du tableau en .csv et mettez-la dans dossier.DB
non
Filtre sur années à rajouter?
oui
Dans GS_MotherCode.r, modifiez la variable filtre.sur.annees
non
Filtre sur espèces à rajouter?
oui
Liste des groupes contenues dans un fichier?
non
Dans GS_MotherCode.r, avec filtre.sur.especes = TRUE, modifiez la valeur des variables taxoF.action, taxoF.utaxo, taxoF.nom
non
oui
Lancement des routines
Manuel
Automatique
Dans GS_MotherCode.r, spécifiez:
sorties.INV = FALSE
sorties.poissons = FALSE
sorties.INV = FALSE
Glissez le fichier GS_MotherCode.r dans R
Dans GS_MotherCode.r, spécifiez:
sorties.INV = TRUE
sorties.poissons = TRUE
sorties.INV = TRUE
Glissez le fichier GS_MotherCode.r dans R
Importez les espèces à filtrer avec la fonction import.filtre.taxo()
Lancez sous R les fonctions désirées, voir:
Run.INV.densite()
Run.INV.biodiv()
Run.poissons.all()
Run.LIT.all()

1 Accès rapide aux descriptions des fonctions R

1.1 Invertébrés

Indices de biodiversité: Shannon, Piélou et Margalef, richesse spécifique:

```
inv.biodiv(AS="pas de filtre",qunit="St",wC="all",save=FALSE)
inv.biodiv.geom(AS="A",save=FALSE)
inv.sprich.tbl(AS="A",grtax="Groupe",save=FALSE)
sprich.by.aggrtaxo(AS="A", grtax="Groupe", save=FALSE)
inv.RichSpecifique(AS="A", aj.impact=FALSE, aggr="St")
```

Densité moyenne sur divers facteurs (transect/station/géomorphologie), avec l'option de conserver les 10 espèces les plus abondantes seulement (tableaux et graphiques):

```
inv.dens.tbl(AS="A", grtax="G_Sp", smpl.unit="St", wZeroAll=FALSE, save=FALSE)
inv.dens.geom(AS="A", aj.impact=FALSE, spttcampagnes=FALSE, save=FALSE)
inv.graph.TS(AS="A", wtype="allsp", wff="Groupe", top10year="", save=TRUE)
inv.graph.TS.top10(AS="A", wff="Groupe", top10year="all", save=TRUE)
```

1.2 Poissons

Données brutes reformattées:

```
poissons.tableau.brut(save=FALSE)
BioDens.sp.poissons()
```

Densité, biomasse, richesse spécifique, etc. par catégorie d'espèce:

```
poissons.ts1(AS="A",save=FALSE)
```

Densité moyenne (par station, année, toutes stations, toutes années):

```
poissons.ts2(AS="A",save=FALSE)
```

Graphiques de séries temporelles biomasse, densité, etc.

```
poissons.p3(quel.graph="all",save=FALSE)
```

1.3 LIT

Données brutes reformattées, couvertures moyennes, etc.:

```
LIT.tableau.brut(save=FALSE, AS="pas de filtre")
LIT.resume(yy=2011, ff="Coraux_Gen", AS="A", save=FALSE)
```

Couverture moyenne par facteurs divers, différence de couvertures entre zones impactées et non-impactées:

```
LIT.ts1(AS="A")
LIT.ts2(AS="A")
```

Graphiques en barres des couvertures moyennes par substrat:

```
LIT.bp1(yy=2011, ff2="Coraux_Gen", AS="A")
```

2 Survol

2.1 Réglages initiaux

Pour lancer le code et produire tous les tableaux/graphiques, commencez par créer deux dossiers à l'emplacement de votre choix:

1. un dossier où les analyses R sont faites [`dossier.R`], et qui contiendra les fichiers R requis par les analyses
2. un dossier où les dernières versions des bases de données sont sauvegardées `[dossier.DB]` (ce dossier doit être synchronisé avec DropBox)

Une fois ces dossiers créés, placez dans le fichier `dossier.R` tous les fichiers .r contenus dans le dossier `cw-ginger-soproner`, disponible en format ZIP à l'adresse suivante: `https://github.com/lauratboyer/cw-ginger-soproner/archive/master.zip`. Ce site contient les dernières versions du code ainsi qu'une historique de toutes les modifications effectuées depuis Janvier 2013.

Ouvrez le fichier `GS_MotherCode.r` [1] et définissez la valeur des objets `dossier.R` et `dossier.DB` en fonction de l'emplacement choisi pour ces dossiers. Pour voir un exemple du format accepté par R pour la définition de l'emplacement de dossiers, tapez dans la console R:

```
> getwd() # montre le working directory
```

Le symbole `#` dénote le début d'un commentaire non-exécuté par R.

Note: Lorsque non-existants, une fonction dans `GS_MotherCode.r` crée automatiquement les trois dossiers suivants dans le `dossier.R`: *Tableaux*, *Graphiques* and *Data*.

2.2 Incorporation de nouvelles versions des bases de données

Les fichiers sont lus par R en format .csv car c'est plus rapide et ne néccessite pas l'installation de librairies additionelles sous R. Pour sauvegarder les tableaux dans le format approprié:

1. ouvrir le tableau dans Excel, aller dans "Enregistrer sous..."
2. sélectionner dans le menu déroulant l'option CSV – attention de sélectionner la *première* option et non CSV - MAC. Si l'ordinateur est un MAC, utiliser l'option CSV - DOS
3. utiliser les noms de fichiers définis au préalable (voir Tableau 1)
4. mettre la nouvelle version du tableau (en .csv) dans le fichier `dossier.DB` défini ci-haut (et ôter la vieille version)

Note 1: Pour s'assurer que la nouvelle version du tableau a bien été incluse lorsque les codes sont lancés, vérifiez que la date imprimée dans la console aprés "Tableaux importés:" correspond bien à la date de création du nouveau fichier pour le tableau .csv en question. Par exemple si une nouvelle version de `Bioeco.csv` est chargée, la date imprimée aura le format suivant:

```
   [1] "Tableaux importés"
$bioeco
[1] "2013-08-01 01:00:00 PST"
```

si le fichier .csv avait initialement était crée le 1er août 2013 à 1 heure.

[1] Un fichier .r peut être modifié sous NotePad ou l'éditeur compris avec R. Si vous programmez plus souvent, *TINN-R* et *R-Studio* sont deux alternatives disponibles gratuitement. Ces logiciels permettent non seulement d'éditer sous R, mais aussi d'interagir plus facilement avec la console.

Tableau 1: Noms de fichiers à utiliser pour la sauvegarde des tableaux en format CSV

Données d'échantillonage invertébrés	Invertebres.csv
Données d'échantillonage poissons	Poissons.csv
Contexte biologie/écologie des poissons	Bioeco.csv
Données de couverture LIT	LIT.csv
Données typologie LIT	Typo_LIT.csv
Variables explicatives transects	Facteurs.csv
Périodes BACIP	Periode_BACIP_Campagnes.csv

Note 2: Pour recharger et reformatter les tableaux manuellement (une fois que `GS_MotherCode.r`a été lançé), utilisez la fonction `import.tableaux()`:

```
> import.tableaux()
```

2.3 Routine de lancement des sorties de base

Dans le fichier `GS_MotherCode.r`, s'assurez que les variables sont bien définies dans la première section. Notamment, les variables `sorties.INV`, `sorties.LIT` et `sorties.POISSONS` devraient avoir la valeur `TRUE` si vous voulez produire les sorties associées, ou `FALSE` dans le cas opposé. La variable `filtre.sur.especes` contrôle les espèces incluses dans l'analyse: lorsque sa valeur est `FALSE` toutes les espèces sont incluses (voir prochaine section).

Vérifiez aussi que l'objet `tourner.tout` (défini dans `GS_MotherCode.r`) a la valeur `TRUE` si vous voulez tout tourner immédiatement, ou `FALSE` si vous voulez charger les fonctions sous R et tourner certains aspects de l'analyse manuellement.
Ouvrez R [2], et chargez le fichier `GS_MotherCode.r` avec la commande `source()`. En début de session seulement, donnez l'emplacement complet de la fonction `GS_MotherCode.r` (donc remplacez ... par l'emplacement du `dossier.R`).

```
> source(".../GS_KNS_MotherCode.r")
```

Note: La première fois que vous lancerez le code, il vous faudra installer certaines librairies R (vous aurez comme message d'erreur: `there is no package called ...`).

Important!! Vous devez charger le fichier `GS_MotherCode.r` au moins une fois en début de session, car ce fichier installe toutes les fonctions nécéssaires à la production des sorties et défini plusieurs variables utilisées dans l'analyse.

2.4 Rajouter un filtre sur les années

Le filtre sur les années est contenu dans l'objet `filtre.sur.annees` défini dans `GS_MotherCode.r`. Pour modifer ce filtre, modifier la valeur de l'objet directement dans `GS_MotherCode.r`avant de lancer `GS_MotherCode.r`sous R. Alternativement vous pouvez définir la valeur de `filtre.sur.annees` directement dans la console pour faire des changements manuellement. Attention si `GS_MotherCode.r`est re-sourcé, la valeur de `filtre.sur.annees`

[2]Une courte introduction aux concepts de bases de R peut se trouver dans les Sections 1 et 2 du document *SimpleR*, disponible à l'adresse suivante: `cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf` (voir aussi les autres sections qui survolent plusieurs applications statistiques).

présente dans le fichier prendra précédence. A tout moment vous pouvez tapez `filtre.sur.annees` dans la console pour voir sa valeur.

2.5 Rajouter un filtre sur les espèces

Si vous désirez inclure ou exclure de l'analyse certains groupes taxonomiques, mettez l'option `filtre.sur.especes = TRUE`. Lorsque `GS_MotherCode.r` est lancé, R vous demandera automatiquement de spécifier les valeurs du filtre sur espèces (voir exemple ci-bas). Pour changer ces valeurs manuellement, vous pouvez également lancez vous même la fonction:

```
> filtre.especes("oui")

   Inclure ou exclure?
      Inclure

   Unité taxomique?  (Groupe/Sous-Groupe/Famille/Genre/Espece)
      Grp2

   Nom?
      Crustaces, Mollusques, Echinodermes
```

Attention: lorsque `filtre.sur.especes = FALSE`, lancer `GS_KNS_MotherCode.r` ôtera automatiquement le filtre enregistré au préalable. Pour ôter *temporairement* ce comportement, vous pouvez rajoutez un # devant la ligne qui exécute cette fonction, en prenant bien soin de l'ôter avant de quitter R pour qu'elle soit active lors de la prochaine session.

```
#if(!filtre.sur.especes){filtre.especes()}else{filtre.especes("oui")}
```

Pour voir la valeur des filtres présentement enregistrée, tapez `voir.filtre.taxo()` dans la console R.

2.6 Modifier les requêtes lancées par défaut

Le fichier `GS_CodesInvertebres_Launch.r` doit être édité pour modifier les requêtes lancées par défaut. Par exemple les arguments des fonctions spécifiées pourraient être changés selon les requêtes voulues.

3 Description des fonctions pour invertébrés, poissons et LIT

3.1 Invertébrés

`inv.biodiv(AS="pas de filtre",qunit="St",wC="all",save=FALSE)`:
Inv_IndexBiodivPar + qunit + filtre station + filtre taxo
Tableau des indices de biodiversité (moyenne et écart type) et richesse spécifique (via `inv.RichSpecifique()`) par station (valeur par défaut, peut être changée en modifiant l'argument `qunit`). Le tableau produit est utilisé par d'autres fonctions, entre autres `inv.biodiv.geom()`.

`inv.biodiv.geom(AS="A",save=FALSE)`:
Inv_IndexBiodiv + géomorphologie/impact + filtre station + filtre taxo
Tableau résumé des moyennes (et ET) sur station par géomorphologie, et par géomorphologie/type d'impact.

`inv.sprich.tbl(AS="A",grtax="Groupe",save=FALSE)`: *Inv_NumEspeceParGeomorph + groupe taxo + filtre station + filtre taxo*
Tableau du nombre d'espèces pour chacun des membres de l'aggrégation taxonomique spécifiée (`grtax`),

proportion des espèces contenues dans chaque membre, et nombre total d'espèces observées par campagne/géomorphologie.

`sprich.by.aggrtaxo(AS="A", grtax="Groupe", save=FALSE)`:
Inv_NumEspeceParAggrTaxon + groupe taxo + filtre station + filtre taxo
Tableau de la richesse spécifique (nombre d'espèces) pour chaque membre du l'aggrégation taxonomique choisie (`grtax`). Par exemple si `grtax = "S_Groupe"`, le tableau contiendra pour chaque station le nombre d'espèce pour chacun des sous-groupes taxonomiques observés sur les stations contenues dans le filtre.

`inv.RichSpecifique(AS="A", aj.impact=FALSE, aggr="St")`:
Non sauvegardé individuellement **Tableau** interne non sauvegardé mais utilisé en information complémentaire pour d'autres tableaux (`inv.biodiv()`, `inv.biodiv.geom()`). Richesse spécifique par niveau taxonomique (nombre de genres, familles, sous-groupes, groupes) par station ou transect, avec option d'ajouter la moyenne (ET) de la RS par géomorphologie (`aggr="geom"`) et zone d'impact (`aj.impact = TRUE`).

`inv.dens.tbl(AS="A", smpl.unit="St", grtax="G_Sp", wZeroAll=FALSE, save=FALSE)`:
Inv_DensitePar_ + smpl.unit + filtre station + groupe taxo + filtre taxo
Tableau des densités moyennes (et SD) par station (ou transect, si `smpl.unit = 'T'`) pour une une unité taxonomique donnée (`grtax`), avec l'option de rajouter dans le tableau final les densités nulles sur les stations/campagne non-occupées (`wZeroAll`).

`inv.dens.geom(AS="A", aj.impact=FALSE, spttcampagnes=FALSE, save=FALSE)`:
Inv_DensityTS_Geomorpho + (Impact) + unité taxonomique + filtre stations + filtre taxonomique + type de top10
Tableaux des densités moyennes (et SD) par espèce, par sous-groupe et par groupe, pour chaque géomorphologie (et zone d'impact si `aj.impact = TRUE`). Les 10 espèces les plus abondantes durant la première, la dernière, et toutes les années de la série temporelle sont sélectionnées. Cette sélection est faite par groupe et par-sous groupe.

`inv.graph.TS(AS="A", wtype="allsp", wff="Groupe", top10year="", save=TRUE)`:
InvDensMoyTS_ + filtre stations + type calcul top10 + année top 10 + niveau taxonomique + géomorpho
Graphiques des séries temporelles de densité (et SD) par unité taxonomique (`wff`) (un graphique par géomorphologie), avec option d'utiliser toutes les espèces (`wtype = 'allsp'`) ou les dix espèces les plus abondantes en 2006 ou sur toute la série temporelle (`wtype = 'top10', top10year = 2006` ou `wtype = 'top10', top10year = 'all'`).

`inv.graph.TS.top10(AS="A", wff="Groupe", top10year="all", save=TRUE)`:
InvDensMoyTS_ + filtre stations + top10 + top10year + niveau taxo + géomorpho + nom membre niveau taxo

Graphiques des séries temporelles de densité (et SD) des 10 espèces les plus abondantes pour chaque membre de l'unité taxonomique (`wff`) (un graphique par géomorphologie et membre de l'unité taxonomique). L'option `top10year` contrôle la selection des dix espèces les plus abondantes (en 20XX, `top10year = 20XX`) ou sur toute la série temporelle (`top10year = all`).

3.2 Poissons

`poissons.tableau.brut(save=FALSE)`:
GS_Poissons_TableauDonneesBrutes_ + date + .csv
Reformattage du **tableau** de données brutes initiales en incluant les valeurs de densité et de biomasse pour chaque observation.

`BioDens.sp.poissons()`:
Tableau préparatoire (non sauvegardé) calculant les densités et biomasses par campagne, transect et espèce en se basant sur l'objet `ds.calc` produit par la fonction `poissons.tableau.brut()`. Le tableau produit par cette fonction (nommé `BDtable` dans l'environnement R) est utilisé par les fonctions `TS1.poissons()` et `TS2.poissons()`.

`poissons.ts1(AS="A",save=FALSE)`:
GS_Poissons_TS1_ + filtre + date + .csv
Tableau synthèse contenant la moyenne et l'écart type pour la densité, la biomasse, la richesse spécifique et la taille moyenne, calculés pour les catégories suivantes: toutes espèces confondues, commerciales, carnivores, herbivores, piscivores, planctonophages, sédentaires, territoriales, mobiles, très mobiles et ciblées (ou non) par la pêche en Nouvelle-Calédonie.

`poissons.ts2(AS="A",save=FALSE)`:
GS_Poissons_TS2_ + filtre + date + .csv
`Tableau` synthèse contenant pour chaque espèce la densité moyenne sur toutes les années, pour chaque année, sur toutes les stations et pour chaque station.

`poissons.p3(quel.graph="all",save=FALSE)`:
GS_PoissonsST_ [valeur représentée].[catégorie de poissons] + date + .pdf
Cette fonction produit des **graphiques** représentant des séries temporelles de densité, biomasse ou richesse spécifique pour les catégories de poissons définies dans l'objet `graph.key$df.id`. Par défaut le filtre annuel est appliqué. Cette fonction est seulement executée dans le code `GS_CodePoissons_Launch.r` si la valeur de l'objet `graph.poissons` est changée à `TRUE`.

3.3 LIT

`LIT.tableau.brut(save=FALSE, AS="pas de filtre")`:
GS_LIT_TableauBrut_ + filtre + date + .csv
Tableau formatté des données de couverture brutes (une rangée par campagne/transect) – ce tableau est sauvegardé dans l'environnement R sous `LITbrut` et est utilisé par les autres fonctions LIT.

`LIT.resume(yy=2011, ff="Coraux_Gen", AS="A", save=FALSE)`:
GS_LIT_Tableau1A_ + code catégorie de coraux + filtre + year + date + .csv
Tableau des couvertures moyennes (et SE) pour chaque géomorphologie selon l'année et la catégorie de substrat spécifiées (par défaut, 2011 et types de substrat généraux).

`LIT.ts1(AS="A")`:
Tableau 1: *GS_LIT_SerieTempTable_ + filtre + _GeoMorphImpact_ + date + .csv*
Tableau 2: *GS_LIT_SerieTempTable_ + filtre + _GeoMorphImpact_DiffCouv_ + date + .csv*
Figures: *GS_LIT_SerieTempTable_ + filtre + _GeoMorphImpact_ + type de coraux + géomorphologie + date + .pdf*
Deux tableaux sont produits. Le premier contient la moyenne (et SE) des couvertures pour chaque campagne, géomorphologie et zone d'impact, le deuxième contient la différence entre les couvertures par géomorphologie des zones impactées et non-impactées. Des **graphiques** sont également produits et comparent pour chaque géomorphologie et catégorie de substrat les tendances de couvertures moyennes par campagne entre les zones impactées et non-impactées.

`LIT.ts2(AS="A")`:
Tableau: *GS_LIT_SerieTempTable_ + filtre + _GeoMorphImpact_bySt_ + date + .csv*
Figure: *GS_LIT_SerieTemp_ + filtre + _GeoMorphImpact_bySt_ + type de coraux + géomorphogie + catégorie d'impact + date + .pdf*
Tableau des couvertures moyennes (et SE) pour chaque type de substrat par station et campagne.

Cette fonction produit aussi des **graphiques** comparant la couverture moyenne de chaque type de substrat entre les stations de la même géomorphologie et zone d'impact.

`LIT.bp1(yy=2011, ff2="Coraux_Gen", AS="A")`:
Nom de fichier: *GS_LIT_Hist1_ + year + code catégorie de coraux + .pdf*
Graphiques en barre comparant la couverture moyenne (et SE) à l'intérieur d'une catégorie de substrat et pour l'année spécifiée.

Note: les catégories de substrats sont définies dans l'objet `coraux.fig`, créé par la fonction `prep.analyse()`.

4 Concepts utiles sous R

Fonctions

Le terme *fonction* sous R dénomme un objet contenant une ou plusieurs lignes de codes qui performent une tâche spécifiée par le programmeur (sous Excel on appellerait ça une macro). Une fonction est très utile lorsqu'on veut répéter les mêmes lignes de code en ne changeant qu'un seul aspect à chaque fois, ou lorsqu'on veut pouvoir lancer une analyse complexe en ne tapant qu'une seule commande. Par exemple si je veux faire le même graphique pour les Crustacés, Mollusques et Échinodermes, je pourrais taper les même ligne de codes 3 fois en changeant les termes 'Crustaces', Mollusques', etc., ou bien je pourrais écrire une fonction qui prend en *argument* le groupe (taxonomique). Par exemple:

```
> fonction.exemple <- function(x) x * 2
> fonction.exemple(4)

[1] 8
```

`function()` est une fonction de base R utilisée pour indiquer qu'on défini un objet de type "function", qu'il s'appelle "fonction.exemple", qu'il prend en argument "x" et qu'il multiplie cet argument par la valeur 2.

```
> fonction.exemple()

Error in x * 2 : 'x' is missing
```

Ici "x" n'a pas de valeur par défaut, donc R retourne un message d'erreur. Lorsqu'on définit une valeur par défaut pour `x` (e.g. `x=3`, voir ci-bas), lancer `fonction.exemple()` ne retourne plus de message d'erreur, mais bien la valeur 6.

```
> fonction.exemple <- function(x=3) x * 2
> fonction.exemple()

[1] 6
```

Pour voir les valeurs par défaut définies dans les fonctions, vérifiez si les arguments sont suivis par un =, ou pas. Si vous lancez une fonction et ne définissez pas un des arguments, R prendra la valeur définie par défaut ou donnera un message d'erreur (si absente). A faire attention, donc: lorsque vous lancez des fonctions manuellement il faut s'assurer que les valeurs définies correspondent bien à ce que vous voulez produire.

5 Visualiser un tableau sous R

De plus, pour visualiser un des tableaux produits directement sous R, il suffit de lancer la commande manuellement dans la console comme suit:

...

Pour voir les premières (ou dernières) 6 lignes du tableau, utiliser la commande `head` (ou `tail`):

... et pour voir le tableau au complet vous tapez le nom de l'objet dans la console.

Pour sélectionnez seulement les lignes où la station à la valeur ...

... et si vous vouliez seulement voir les colonnes ...

Les opérateurs logiques sous R sont: == est égal à != n'est pas égal à x & y les deux conditions doivent être respectées x | y l'une ou l'autre des conditions doivent être respectées