

# COLLECTER - TRAITER - REPRESENTER

## Gagnez du temps grâce à l'automatisation



INNOVATION ■ OUTILS ■ PROJETS ■ TERRITOIRES

Simon Georget

**/ La situation**

**/ Comment a-t-on procédé ? Choix et outils**

**/ Pour aboutir à quels résultats ? Et quelles représentations ?**

# I - La situation & les données

Pourquoi en sommes-nous arrivés à penser l'automatisation ?

## / Intermezzo : bureau d'études et de conseils sur les questions de développement durable territorial

- \_ transition énergétique
- \_ changement climatique
- \_ mobilité
- \_ suivi / évaluation / concertation / ...

## / Approche spatiale forte

## / Culture de la donnée

## / Nous réalisons régulièrement des PCAET : Plan Climat Air Energie Territorial

## DONNEES / Nos besoins récurrents pour l'élaboration des diagnostics PCAET et les sources de données

### Les Observatoires Régionaux de l'Énergie et des Gaz à Effet

\_ Bilan distinguant les secteurs : Résidentiel, Industrie, Tertiaire, Agricole, Transport

### → La plus value du BE : comprendre les déterminants

\_ Pour cela, se plonger dans la donnée brute fournie par une multitude d'acteurs. Historique de données.

## DONNEES / Nos besoins récurrents pour l'élaboration des diagnostics PCAET et sources de données

- \_ **Données INSEE** : Base Logement (3 années de référence), Historique (depuis 1968), Equipement, Mobilité (2006, 2011, 2013), Flux domicile-travail 2010, SIRENE
- \_ **Données des opérateurs énergétiques** : Production / Consommation d'énergie (électricité et gaz) à l'échelle de l'IRIS  
Sources biens connus GRDF / ENEDIS / GRTGaz / RTE ... et ~ 170 ELD
- \_ **Données du SDES / SoeS** : Compilation des données opérateurs : conso / prod / livraison.  
Production de chaleur et de froid. Produits pétroliers à l'échelle départementale seulement
- \_ **Données relatives au risques** : Géorisques / Base Gaspar (arr. Catastrophes / risques), BRGM RGA, ICPE
- \_ **Données spatiales** : IGN (adminexpress / IRIS) + croisement de données

## DONNEES / Nos besoins récurrents pour l'élaboration des diagnostics PCAET et sources de données

Récapitulons :

- \_ Collecte multisources ennuyeuse
- \_ La plupart du temps, données au format Excel, fichiers plats
- \_ Manipulation de données très lourde sous Calc / Excel.

On a décidé d'en finir avec ça ....

## II - Un pas vers la lumière

Des briques libres pour automatiser la collecte et préparer les données

## AUTOMATISER / Les briques logicielles utilisées

`#!/bin/bash`

CSVKit

GDAL / OGR

WGET

Panda (py)

PHP

QGIS

GRASS GIS

### Au delà des outils :

- \_ Conception (relativement) flexible de l'outil : inputs, résultats thématiques
- \_ Définir les formats et l'architecture de rendus
- \_ Préparer les données "inertes" pour réduire les temps d'exécution du script

# Bash / Le choix de la simplicité

## Un EPCI == un ensemble de communes

\_ Préparation d'un fichier communes.csv sur lequel je vais itérer

```

1 "Insee";Nom;Gentilé;"Superficie(km2)";"Population(dernière pop. légale)";"Densité(hab./km2)"
2 56130;"Merlevenez(siège)";Merleveneziens;"17,67";3 175 (2014);180
3 56094;Kervignac;Kervignacois;"39,56";6 525 (2014);165
4 56148;Nostang;Nostangais;"15,71";1 505 (2014);96
5 56169;Plouhinec;Plouhinecois;"35,58";5 301 (2014);149
6 56220;Sainte-Hélène;Hélinois;"8,08";1 187 (2014);147

```

## Lire un fichier CSV en bash

```

1 #!/bin/bash
2
3 INPUT="communes.csv"
4 OLDIFS=$IFS
5 # ici, on précise le séparateur
6 IFS="$separateur"
7
8 echo "" >> $INPUT
9
10 [ ! -f $INPUT ] && { echo "$INPUT file not found"; exit 99; }
11 while read -r f1 f2 f3 f4 f5
12 do
13     # on teste si c'est un nombre - avec une regex
14     re='^[0-9]+$'
15     if [[ $f1 =~ $re ]] ; then
16         echo -e "Récupération des données pour la commune ${f1}"
17     fi
18
19 done < $INPUT
20 IFS=$OLDIFS

```

Points de vigilance :

- \_ Connaitre le séparateur
- \_ S'assurer que la dernière ligne soit bien lue

# Bash / Le choix de la simplicité

## Collecter la base geoSIRENE pour un EPCI (Merci cquest !)

```

1 [ ! -f $INPUT ] && { echo "$INPUT file not found"; exit 99; }
2 while read -r f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12
3 do
4     # on teste si c'est un nombre - avec une regex
5     re='^[0-9]+$'
6     if [[ $f1 =~ $re ]]; then
7         echo -e "${cyanclair}${COUNTER} - Récuperation des données cartographiques pour la commune ${f1} ${neutre}"
8
9     # on télécharge le fichier que s'il n'est pas déjà dans le répertoire temporaire
10    if [ -f ${tmpfolder}${f1}.csv ]
11    then
12        echo -e "${jaune}Le fichier ${f1}.csv est déjà présent dans ${tmpfolder}${neutre}"
13    else
14        echo -e "${jaune}Téléchargement du fichier communal Géo SIRENE : ${geosireneURLcomm//COMM//${f1}}${neutre}"
15        echo -e "${rose}"
16        wget ${geosireneURLcomm//COMM//${f1}} -P ${tmpfolder}
17        echo -e "${neutre}"
18    fi
19
20    COUNTER=$((COUNTER + 1))
21 fi
22
23
24 done < $INPUT
25 IFS=$OLDIFS
26
27 # Fusionne l'ensemble des fichiers en excluant la première ligne sauf pour le premier fichier.
28 # https://askubuntu.com/questions/656039/concatenate-multiple-files-without-header
29 awk 'NR==1 {header=$_} FNR==1 && NR!=1 { $_ ~ $header getline; } {print}' ${tmpfolder}*.csv > ${tmpfolder}${outputfilename}
```

\_ Apport de wget / Puissance de la ligne de commande (awk ici)

# Zoom sur CSVKit / La pièce majeure !

## Why csvkit?

Because it makes your life easier.

Convert Excel to CSV:

```
in2csv data.xls > data.csv
```

Convert JSON to CSV:

```
in2csv data.json > data.csv
```

Print column names:

```
csvcut -n data.csv
```

Select a subset of columns:

```
csvcut -c column_a,column_c data.csv > new.csv
```

Reorder columns:

```
csvcut -c column_c,column_a data.csv > new.csv
```

***csvkit is a suite of command-line tools for converting to and working with CSV, the king of tabular file formats.***

*It is inspired by pdftk, gdal and the original csvcut tool by Joe Germuska and Aaron Bycoffe.*

*If you need to do more complex data analysis than csvkit can handle*

**Plus sur CSVKit :**

<https://csvkit.readthedocs.io/en/1.0.3/>  
<https://github.com/wireservice/csvkit>

## Zoom sur CSVKit / La pièce majeure !

### Quelques exemples appliqués à notre utilisation :

```
csvgrep -d ";" -c 1 -m "$f1" input.csv >> output.csv
```

```
csvgrep -d ";" -e iso-8859-1 -c "CODE" -r "^$f1" input.csv >> output.csv
```

```
csvgrep -d ";" -c 4 -m "$f1" input.csv | tail -n +2 >> output.csv
```

# Zoom sur CSVKit / La pièce majeure de notre expérience

## Quelques fonctionnalités complémentaires :

Generate summary statistics:

```
csvstat data.csv
```

Query with SQL:

```
csvsql --query "select name from data where age > 30" data.csv > new.csv
```

Import into PostgreSQL:

```
csvsql --db postgresql://database --insert data.csv
```

Extract data from PostgreSQL:

```
sql2csv --db postgresql://database --query "select * from data" > new.csv
```

```
19. CMBL
<type 'int'>
Nulls: False
Min: 1
Max: 6
Sum: 113200
Mean: 3.07216326974
Median: 2
Standard Deviation: 1.4341378503
Unique values: 6
5 most frequent values:
 2:    18550
 4:    6883
 3:    5053
 6:    5020
 1:    854
```

- Combustible principal du logement :**
- 1 : Chauffage urbain du logement (France)
  - 2 : Gaz de ville ou de réseau métropolitaine)
  - 3 : Fioul (mazout)
  - 4 : Electricité
  - 5 : Gaz en bouteilles ou en citerne
  - 6 : Autre

## DONNEES : SOURCES & QUALITE / Les bons choix

### Faire des choix relatifs à la redondance des données

La LTECV (Art.179) facilite l'accès aux données provenant des gestionnaires de réseaux de distribution et de transport d'énergie. Ces données concernent la production, le transport, la distribution et la consommation d'énergie (électricité, gaz, produits pétroliers, froid et chaleur).

Le décret n°2016-973 du 18 juillet 2016 relatif à la mise à disposition des personnes publiques de données relatives au transport, à la distribution et à la production d'électricité, de gaz naturel et de biométhane, de produits pétroliers et de chaleur et de froid et l'arrêté du 18 juillet 2016 fixent les modalités de transmission de ces données.

Notamment sur les données de consommation d'énergie :

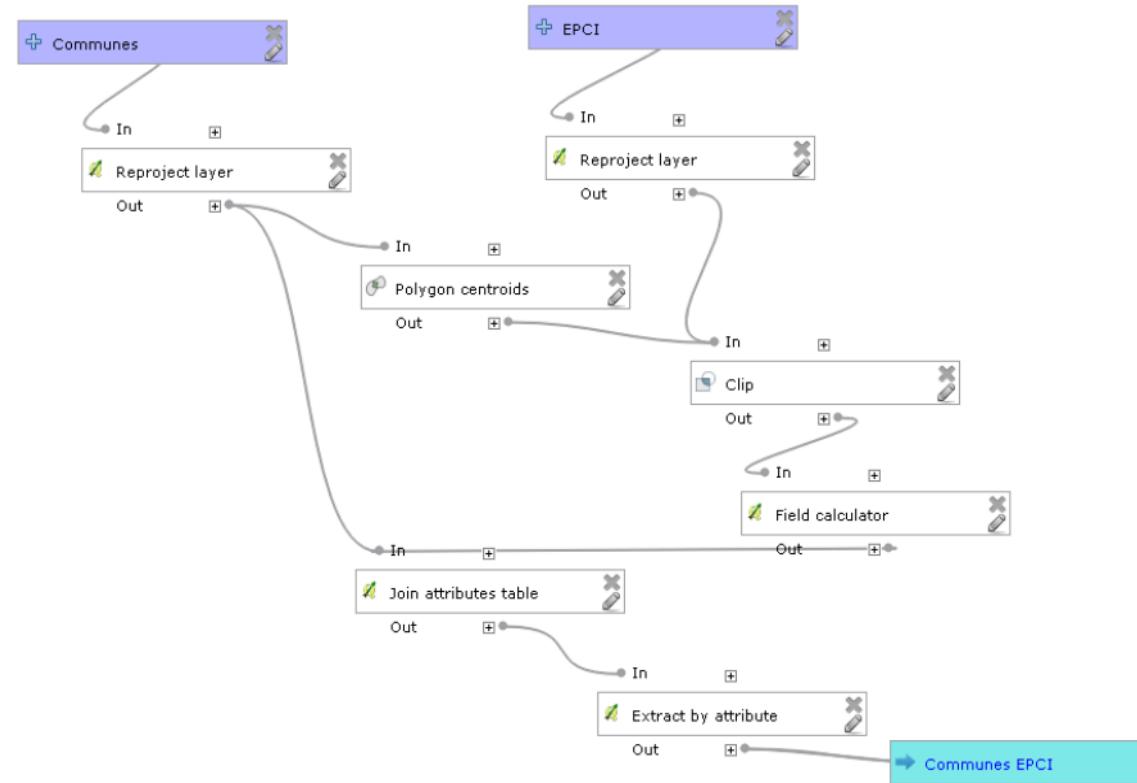
**SoeS / SDES ou opérateurs ?**

## III – Représenter les données

Automatiser la production cartographique

## Générer nos fonds de carte / Le modeleur QGIS ?

- Extraction des communes en fonction des limites d'un EPCI



Boris Mericskay - Chaînes de traitements automatisées - Prise en main du modeleur graphique de QGIS

## Générer nos fonds de carte / La puissance d'OGR

```

45 # Fonds cartographiques - fichiers sources
46 adminexpress="/data/ADMINEXPRESS_1-0__SHP_LAMB93_FXX_2016-09-16/ADMINEXPRESS/1_DONNEES_LIVRAISON_2016-09-16
/ADE_1-0_SHP_LAMB93_FR/COMMUNE.shp"
47 iriscontour="/data/CONTOURS-IRIS_2-1__SHP_LAMB93_FXX_2016-11-10/CONTOURS-IRIS/1_DONNEES_LIVRAISON_2015/CONTOURS-
IRIS_2-1_SHP_LAMB93_FE-2015/CONTOURSIRIS.shp"
48
49 # Nom des fichiers de sortie
50 outputfilename_adminexpress="adminexpress_territoire.shp"
51 outputfilename_iriscontour="iriscontour_territoire.shp"
52 separateur=";"                      # séparateur du fichier CSV !!! A PRÉCISER !!! IMPORTANT

```

### \_ Itération sur le fichier CSV

```

152 # On travaille sur le fichier commune ADMIN EXPRESS
153 ogr2ogr -f "ESRI Shapefile" -sql "SELECT * FROM ${aefilenamenoext} WHERE INSEE_COM='${f1}'"
${tmpfolder}_adminexp_${COUNTER}.shp ${adminexpress}
154
155 # On fusionne les fichiers ADMIN EXPRESS
156 if [ ! -f "${tmpfolder}${outputfilename_adminexpress}" ]
157 then
158     echo -e "... Création du fichier ${tmpfolder}${outputfilename_adminexpress}"
159     ogr2ogr ${tmpfolder}${outputfilename_adminexpress} ${tmpfolder}_adminexp_${COUNTER}.shp -f "esri shapefile"
160 else
161     echo -e "... Fusion de ${tmpfolder}_adminexp_${COUNTER}.shp avec ${tmpfolder}${outputfilename_adminexpress}."
162     ogr2ogr -update -append ${tmpfolder}${outputfilename_adminexpress} ${tmpfolder}_adminexp_${COUNTER}.shp -f "esri
shapefile" -nln ${outputfilename_adminexpress%.*}
163 fi

```

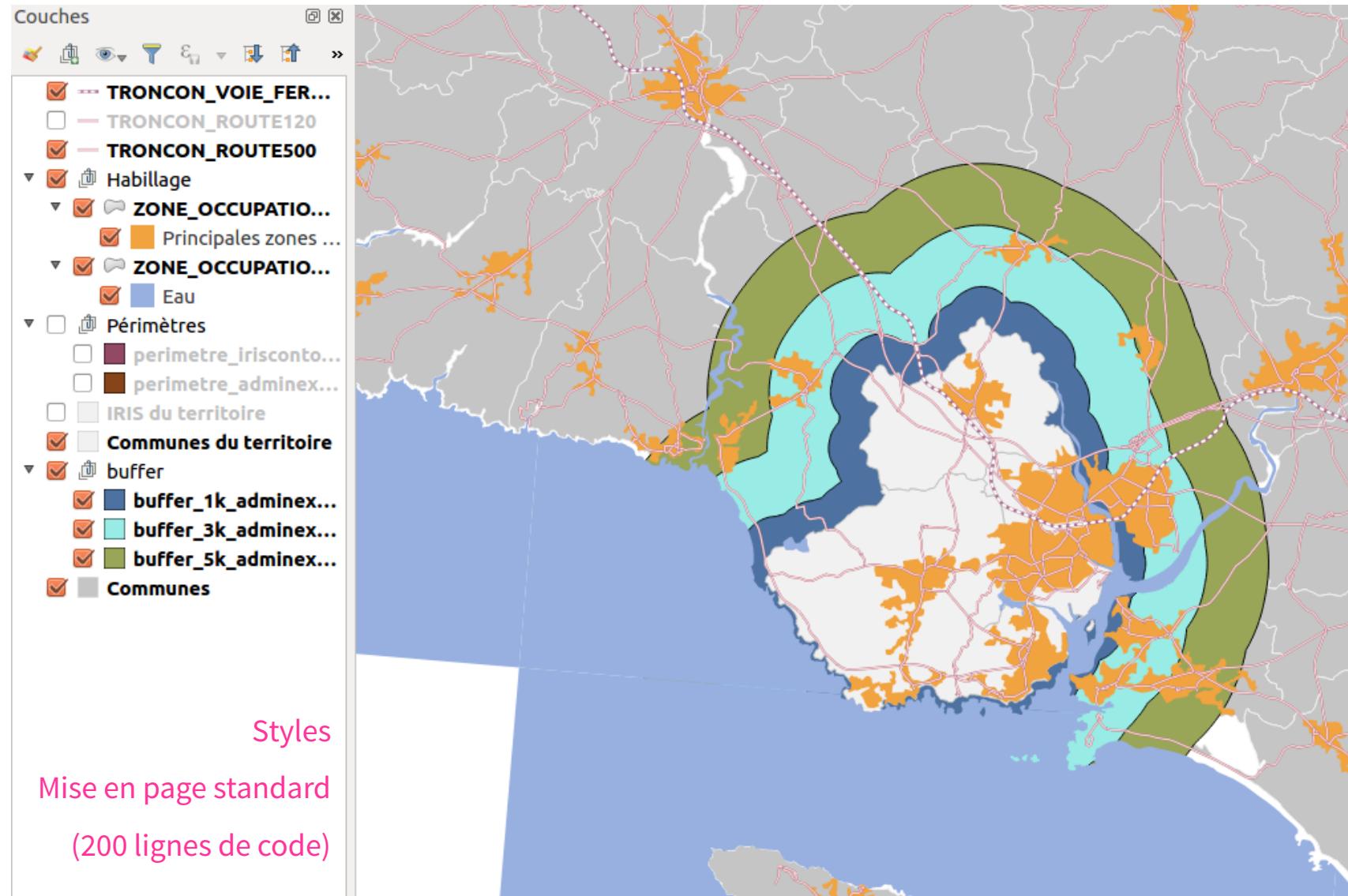
## Générer nos fonds de carte / La puissance d'OGR

\_ Même opération sur la couche CONTOURIRIS

\_ Bonus pour préparer nos cartes et traitements complémentaires

```
208 echo -e "... Création du périmètre unique - couche ${outputfilename_adminexpress}"
209 ogr2ogr -f "ESRI Shapefile" ${tmpfolder}perimetre_${outputfilename_iriscontour} ${tmpfolder}${outputfilename_iriscontour}
-dialect sqlite -sql "select ST_union(Geometry),* from ${outputfilenameadminext%.*}"
210
211 echo -e "... Création des couches tampons : 1km, 3km, 5km"
212 ogr2ogr -sql "SELECT ST_Buffer(A.geometry,1000) AS geometry, A.* FROM perimetre_${outputfilename_adminexpress%.*} A" -dialect
SQLITE ${output}_cartographie/buffer_1k_${outputfilename_adminexpress}
${output}_cartographie/perimetre_${outputfilename_adminexpress}
213
214 ogr2ogr -sql "SELECT ST_Buffer(A.geometry,3000) AS geometry, A.* FROM perimetre_${outputfilename_adminexpress%.*} A" -dialect
SQLITE ${output}_cartographie/buffer_3k_${outputfilename_adminexpress}
${output}_cartographie/perimetre_${outputfilename_adminexpress}
215
216 ogr2ogr -sql "SELECT ST_Buffer(A.geometry,5000) AS geometry, A.* FROM perimetre_${outputfilename_adminexpress%.*} A" -dialect
SQLITE ${output}_cartographie/buffer_5k_${outputfilename_adminexpress}
${output}_cartographie/perimetre_${outputfilename_adminexpress}
217
```

## Générer nos fonds de carte / Nous ouvrons enfin QGIS !



# Générer nos fonds de carte / Nous ouvrons enfin QGIS !

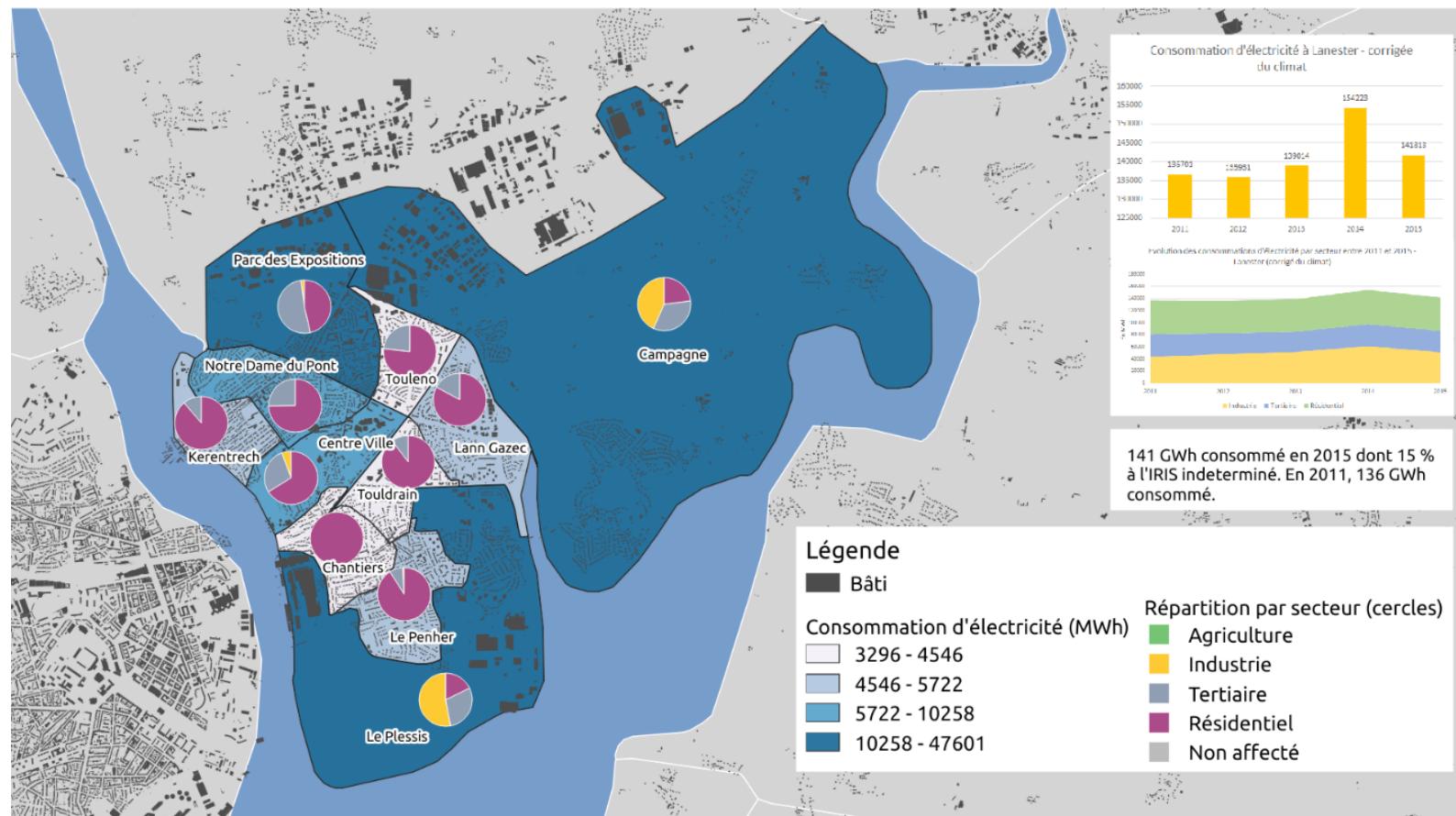
## – Une macro pour zoomer sur la zone concernée

```
1 - def openProject():
2     from qgis.gui import QgsMapCanvas;
3     from qgis.gui import QgsMessageBar;
4     registry = QgsMapLayerRegistry.instance();
5     canvas = iface.mapCanvas();
6     name= 'buffer_5k_adminexpress_territoire';
7     layer = registry.mapLayersByName( ·name · )[0];
8     canvas.setExtent(layer.extent());
9     canvas.refresh();
10    iface.messageBar().pushMessage("Info", ·"Zoom sur le territoire", ·level=QgsMessageBar.INFO);
11    pass
12
```

# Représenter / Prendre le temps pour l'analyse et la création

## Consommation d'électricité en 2015 et répartition par secteur

Une consommation en hausse de 4 % depuis 2011 due à l'industrie (données corrigées du climat).  
Sur la période 2011-2015 : Industrie + 17 %, Tertiaire -5 %, Résidentiel -1 %

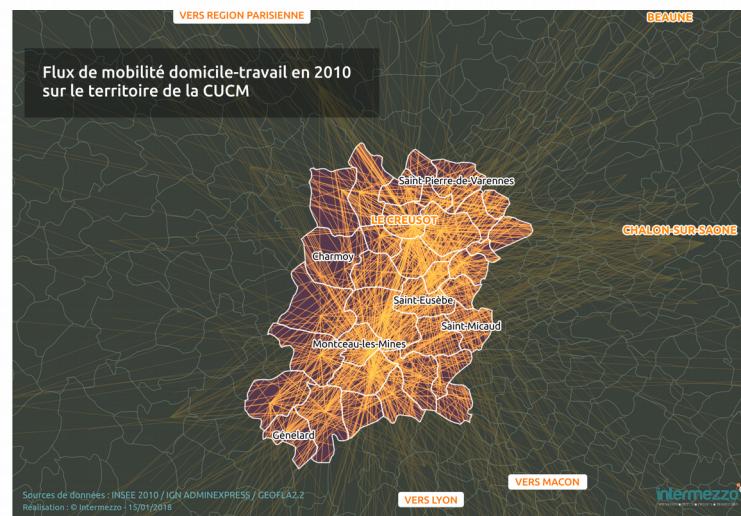
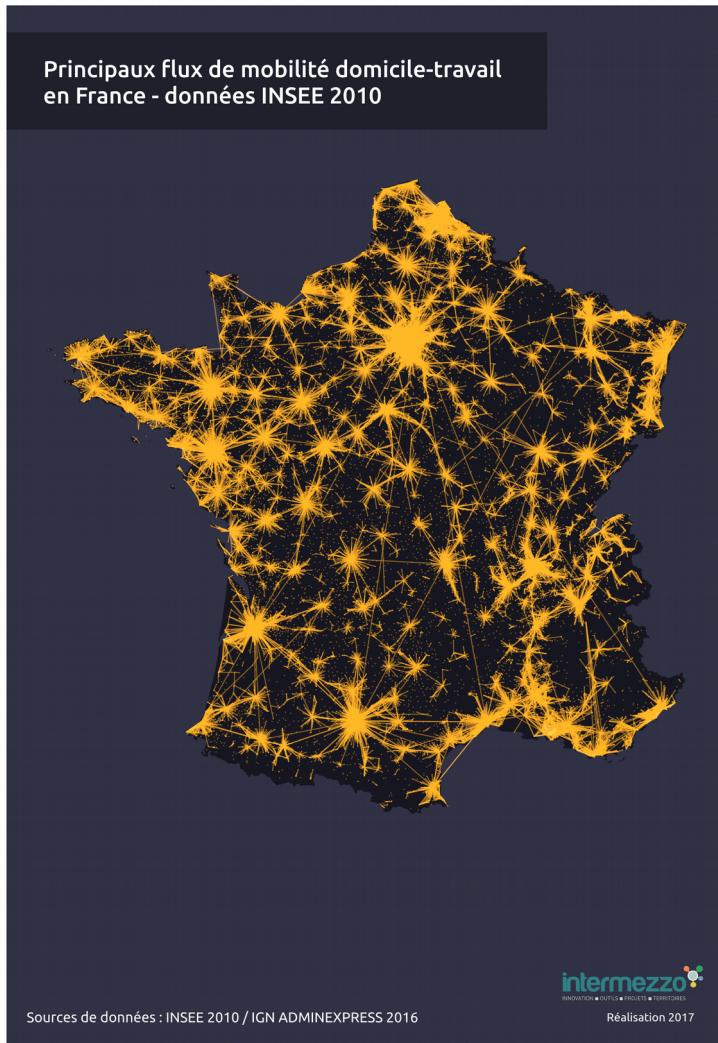


intermezzo

## Représenter / Plus de temps pour l'analyse et la création

### Mapping flows - Alasdair Rae

<http://www.undertheraedar.com/2014/10/flow-mapping-with-qgis.html>

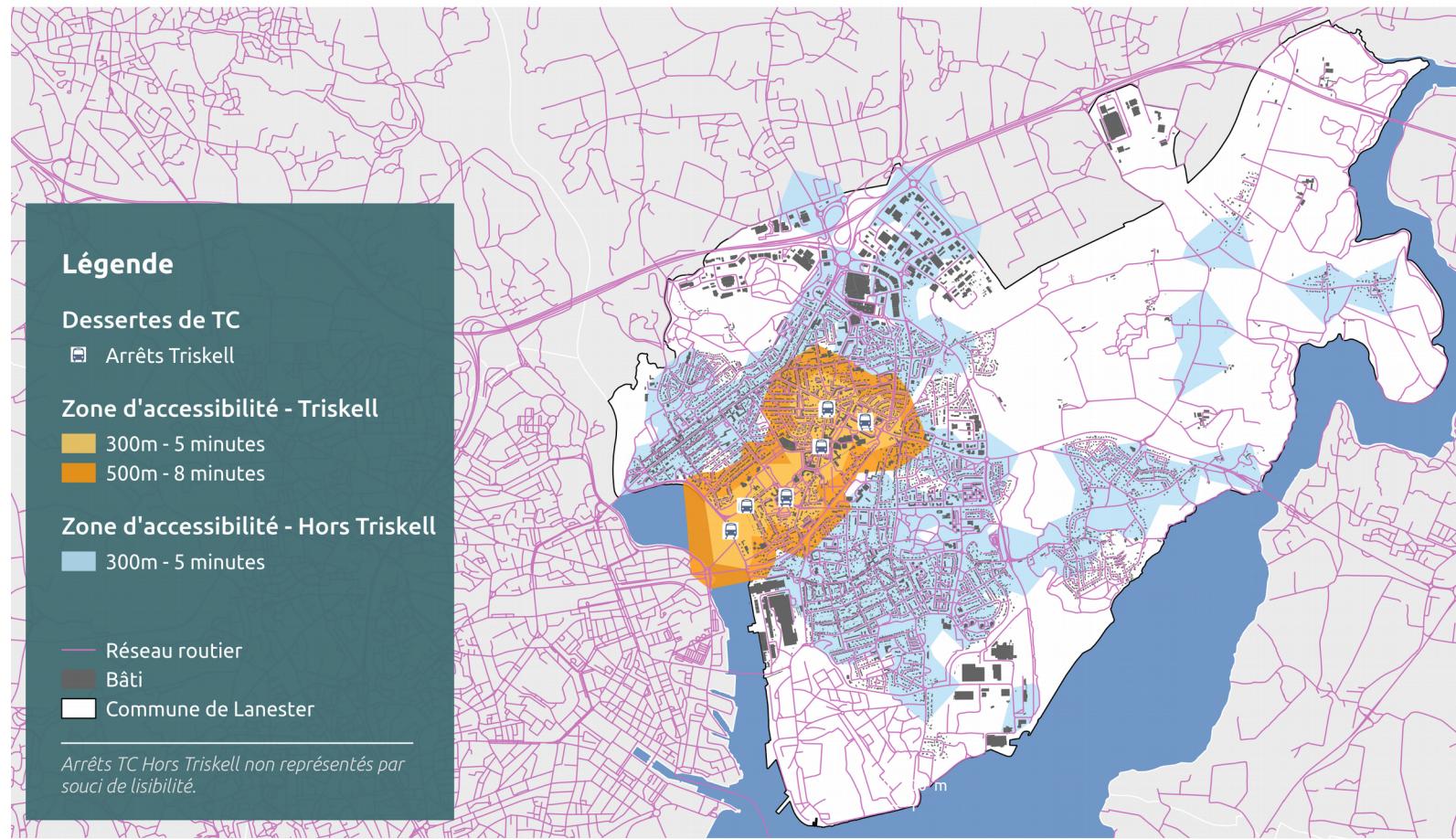


# Représenter / Plus de temps pour l'analyse et la création

## Lanester - Zones d'accessibilités piétonnes aux dessertes de transport en commun

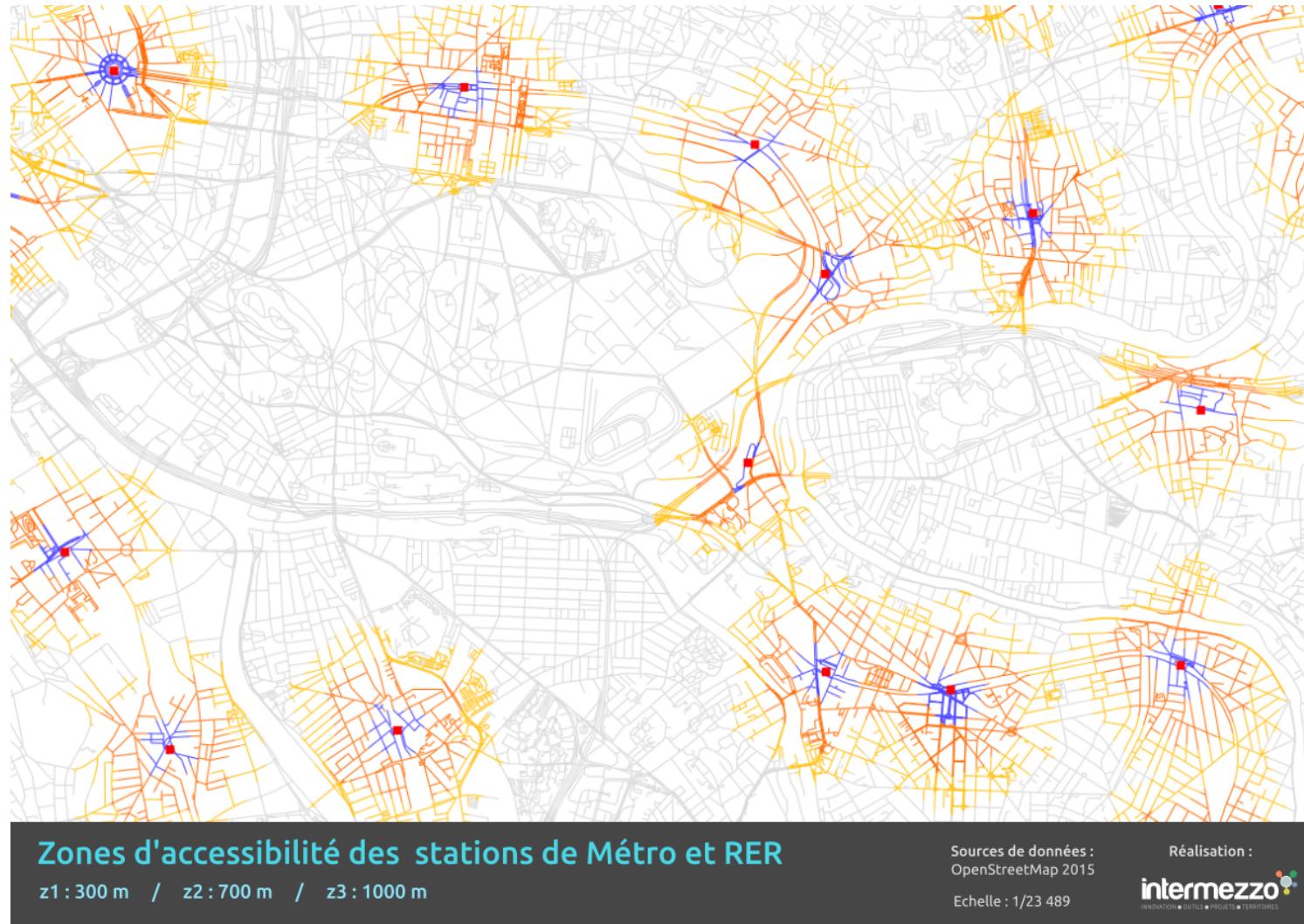
10 % des espaces bâties de Lanester se trouvent dans une **zone d'accessibilité de 300m du Triskell**. Cette part est de **23 %** lorsque la zone d'accessibilité est de **500m**. Plus de 80 % des espaces bâties sont dans une zone d'accessibilité des transports en commun inférieure à 300m.

### ISOCRONES Réalisés à l'aide de GRASS GIS



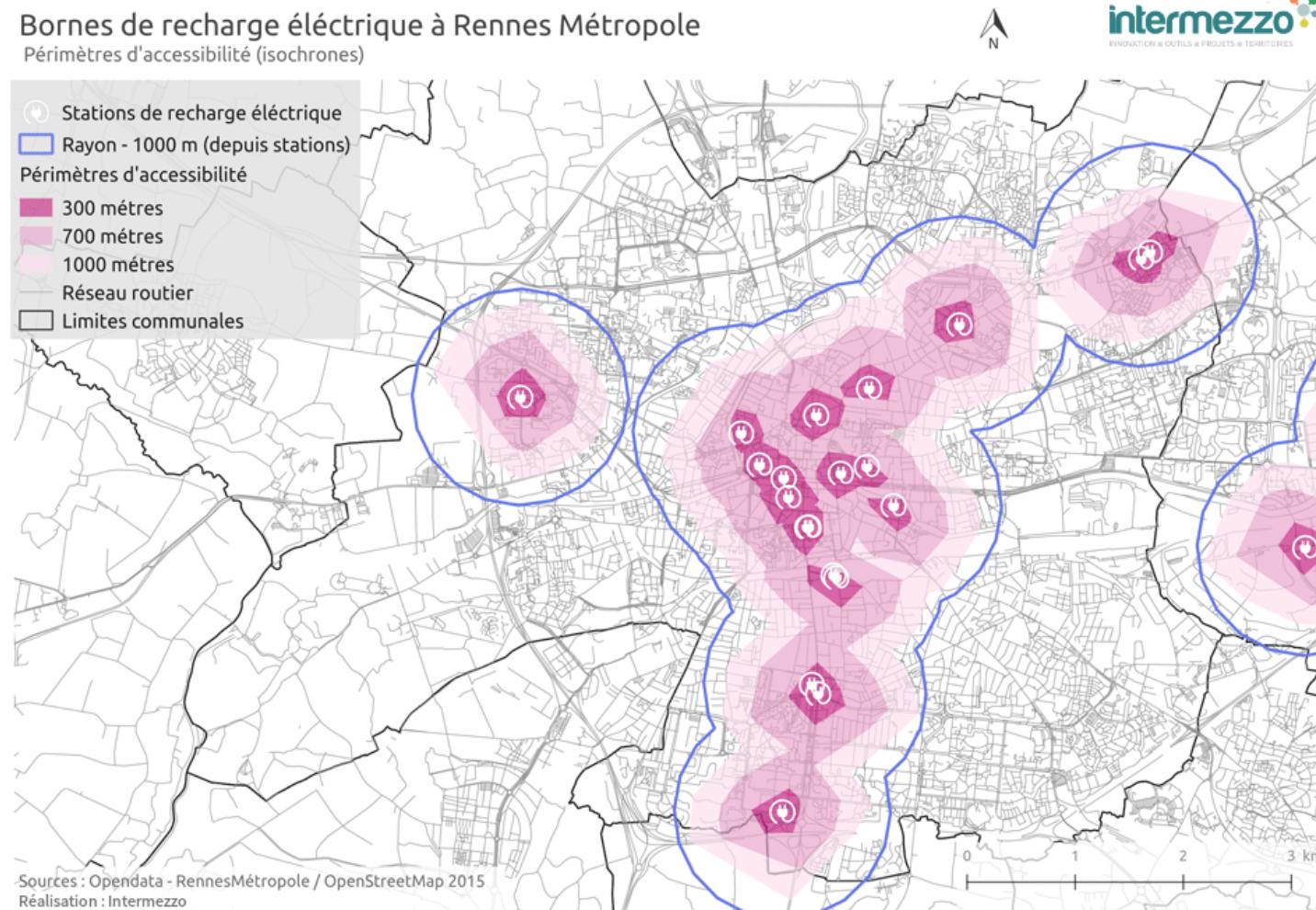
## Représenter / Plus de temps pour l'analyse et la création

**ISOCHRONES**  
Réalisés à l'aide de GRASS GIS



## Représenter / Plus de temps pour l'analyse et la création

### ISOCRONES Réalisés à l'aide de GRASS GIS



## S'il fallait conclure ...

- \_ Des métiers dans lesquels on passe trop de temps à repousser l'automatisation
  - \_ Elle peut se faire tant pour la collecte, le traitement et la représentation
  - \_ Bash est plutôt facile d'accès
  - \_ Gain de temps évident
  - \_ Permet une accession à un public non averti
  - \_ Homogénéité des méthodes (données et sources par exemple)
  - \_ Evolutif
  - \_ On pourrait envisager d'automatiser les traitements stats les plus fréquents
- **Se diriger vers d'autres outils ?**

## Dernier conseil : soyez un esthète du shell !

```
simon@hal:~$ cd Bureau/FOSS4G-fr/presentation-Intermezzo/
simon@hal:~/Bureau/FOSS4G-fr/presentation-Intermezzo$ ./couleur-bash.sh
---Bg---40---41---42---43---44---45---46---47
fg30- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg31- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg32- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg33- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg34- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg35- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg36- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
fg37- SS64 SS64 SS64 SS64 SS64 SS64 SS64 SS64
-- Clear BG --
fg30 SS64
fg31 SS64
fg32 SS64
fg33 SS64
fg34 SS64
fg35 SS64
fg36 SS64
fg37 SS64
```

\_ script disponible sur : <https://ss64.com/bash/echo.html>



INNOVATION ■ OUTILS ■ PROJETS ■ TERRITOIRES

## Merci de votre attention

### Contact

[www.intermezzo-coop.eu](http://www.intermezzo-coop.eu)

 @intermezzoCoop

[simon@intermezzo-coop.eu](mailto:simon@intermezzo-coop.eu)

 @empurb



<https://github.com/simogeo/geostats>