

ATELIER ATLAS

INFOGRAPHIES

RÉALISÉES EN PARALLÈLE

DES CARTES

L'atelier Atlas consistait en la réalisation de cartes et d'infographies en réponse à une commande de France Volontaires. Ce fut un travail très enrichissant puisqu'il m'a permis de maîtriser de nouveaux outils et de découvrir un autre environnement de travail à la Maison des Suds.

Pour ma part je comptais sur le master pour acquérir de nouvelles compétences techniques et j'ai saisi l'occasion pour m'intéresser de plus près à de nouveaux logiciels: R et D3. J'ai essayé de les utiliser pour proposer, à côté des cartes à réaliser, d'autres infographies. En voici le résultat.

Les premières données étaient sous la forme d'un tableau avec en individus les missions réalisées et en colonne différentes caractéristiques de la mission et du volontaire l'ayant effectuée. D'autres tableaux concernaient les chantiers et les accueils de volontaires en France.

Toutes les cellules n'étaient pas renseignées et certaines valeurs apparaissaient plusieurs fois sous des orthographes différentes. J'ai entrepris de réaliser manuellement un appareillage des données. Ce fut long et je n'ai pas terminé mais c'était de toute façon nécessaire pour réaliser certaines cartes et graphiques. J'ai tenté de placer le fichier corrigé sur une plateforme comme Google Docs pour que nous puissions tous apporter des corrections au fur et à mesure mais le tableau était trop volumineux. Au final, chaque groupe a dû intervenir sur la donnée de manière différente selon la carte à réaliser.

```
# Charger la feuille de donnees dans R
vol <- read.csv(file = "env.csv", sep=";", encoding="UTF-8")
library(dplyr)
library(ggplot2)
```

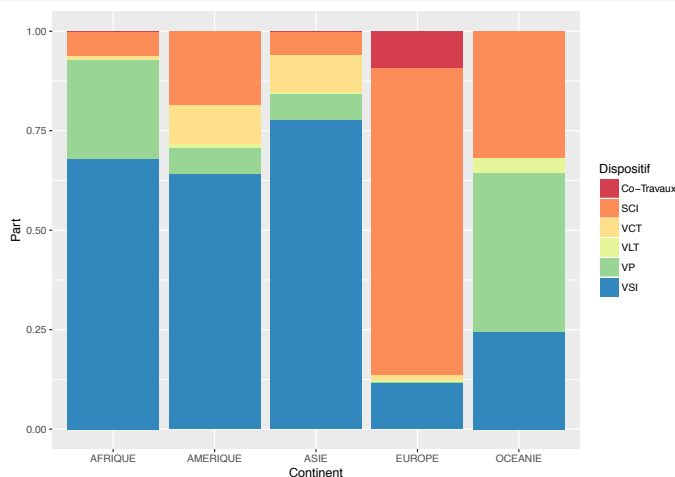
En cours de semestre, les commanditaires ont exprimé le désir de ne compter que les missions à partir de 2009. C'est l'année de création de France Volontaires et de certains dispositifs qui ne voulaient pas être sous-représentés. On peut estimer l'impact de cette décision avec le graphique ci-dessus. Le volume des missions en VSI est de loin le plus important mais la différence est tout de même visible.

À l'aide de ggplot2, on peut facilement réaliser une grille de graphiques. Les graphiques comparés en bas représentent l'évolution du nombre de missions par an. L'Afrique est loin devant.

Les graphiques sur la page de droite montrent la préférence pour un continent d'envoi ou pour un dispositif en fonction de l'âge. La courbe des âges montre un plus grand nombre de 20-30 ans. Le dispositif SCI est responsable du pic juste avant 26 ans. En bas, le cumul des durées des missions sur chaque continent par année montre un investissement moindre des volontaires en Afrique entre 2009 et 2015, tandis que l'Europe et l'Océanie progressent.

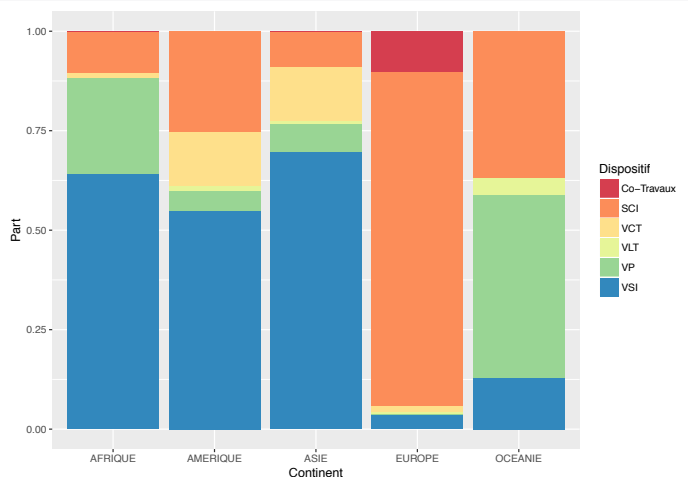
TOUTES LES DONNÉES

```
ggplot(vol, aes(Gdes_regions)) + geom_bar(position="fill",
aes(fill=Dispositif)) + labs(y="Part", x="Continent") +
scale_fill_brewer(palette="Spectral")
```

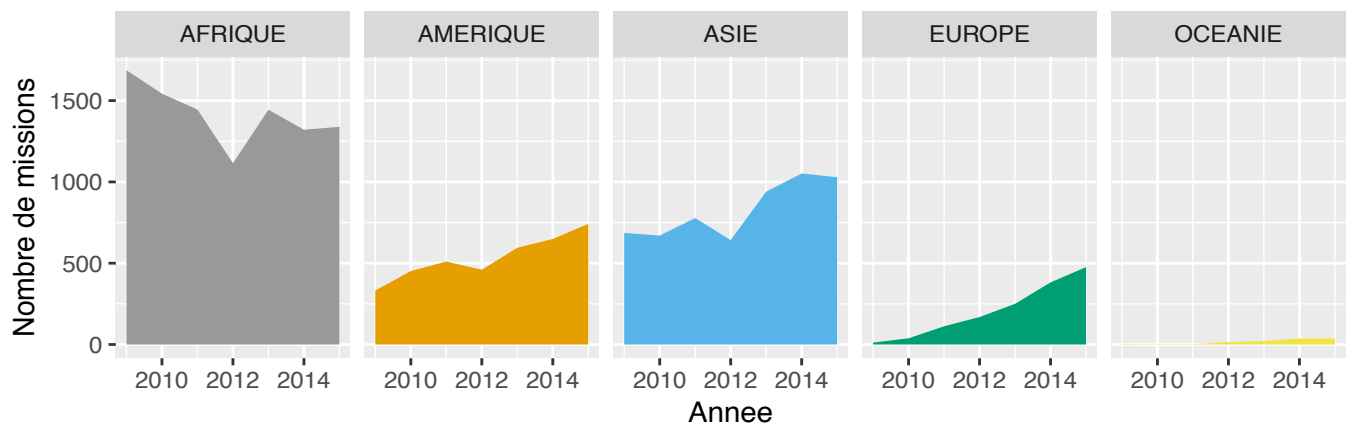


DONNÉES À PARTIR DE 2009

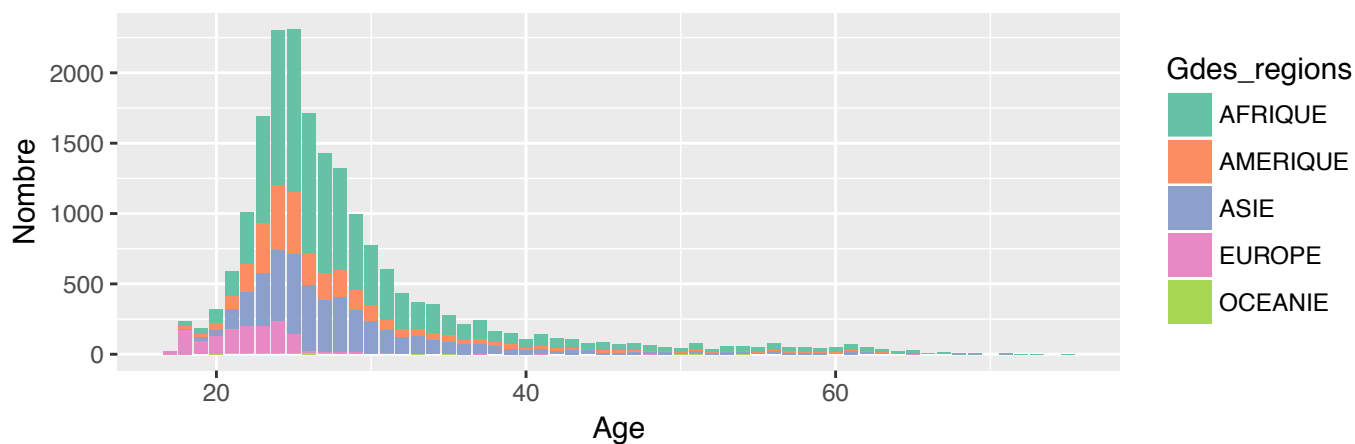
```
ggplot(filter(vol, Annee >= 2009), aes(Gdes_regions))
+ geom_bar(position="fill", aes(fill=Dispositif)) +
labs(y="Part", x="Continent") + scale_fill_brewer(pal-
ette='Spectral')
```



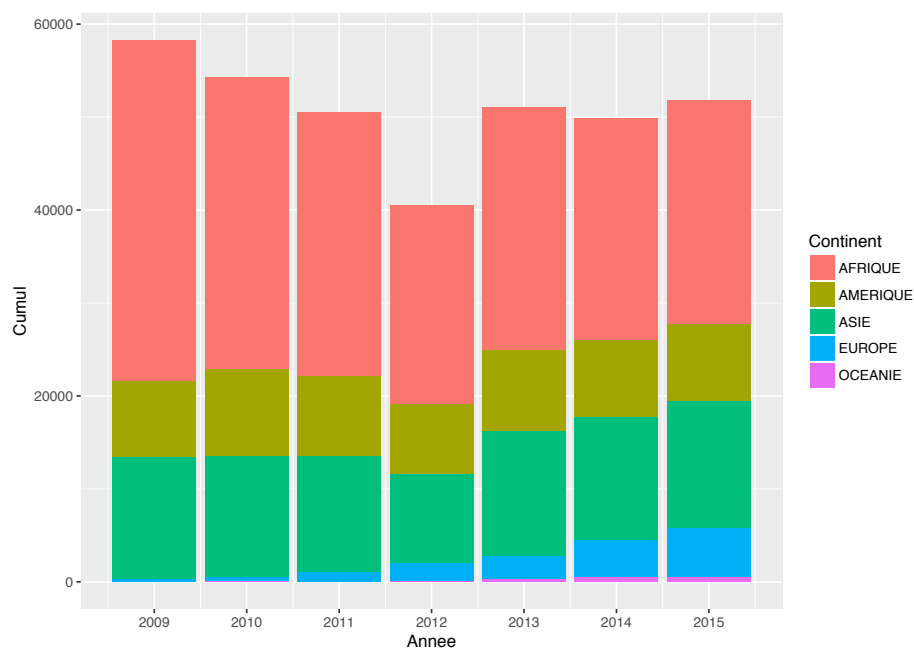
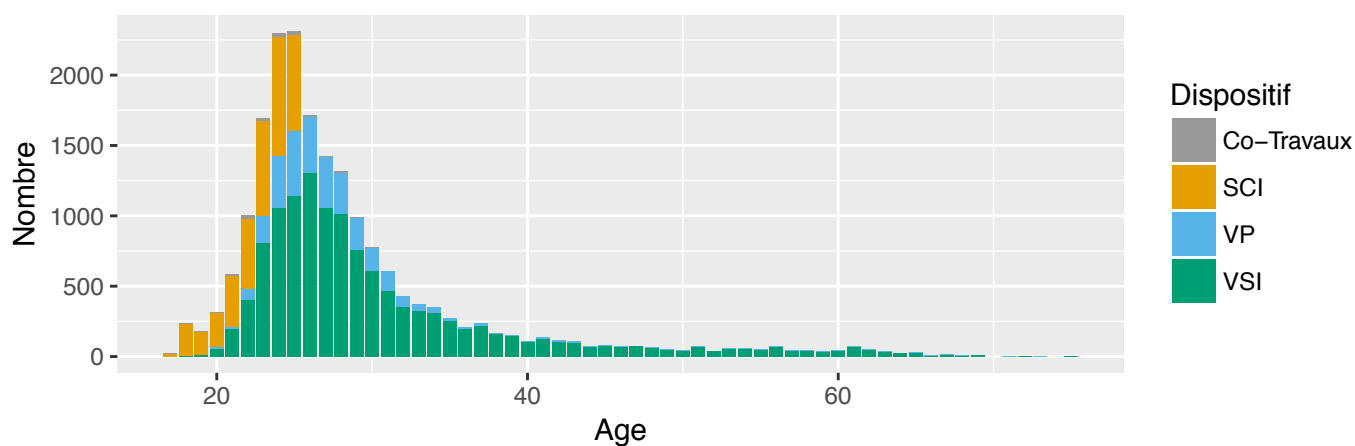
```
vol0915 <- subset(vol, Annee>=2009)
ggplot(vol0915, aes(Annee)) + geom_area(stat="count",aes(fill=Gdes_regions)) + facet_grid(.~Gdes_regions) + labs(y="Nombre de missions", x="Annee", fill="Continent d'envoi") + theme(legend.position="none") + scale_fill_manual(values=cbPalette)
```



```
ggplot(filter(vol0915), aes(Age)) + geom_bar(aes(fill=Gdes_regions)) + scale_fill_brewer(palette="Set2") + labs(y="Nombre")
```



```
ggplot(filter(vol0915), aes(Age)) + geom_bar(aes(fill=Dispositif)) + scale_fill_manual(values=cbPalette) + labs(y="Nombre")
```



```
head(durees)
  Annee Continent      Cumul
1  2009   AFRIQUE 36644.53333
2  2009  AMERIQUE 8116.70000
3  2009    ASIE 13142.46667
4  2009   EUROPE  255.73333
5  2009  OCEANIE   72.93333
```

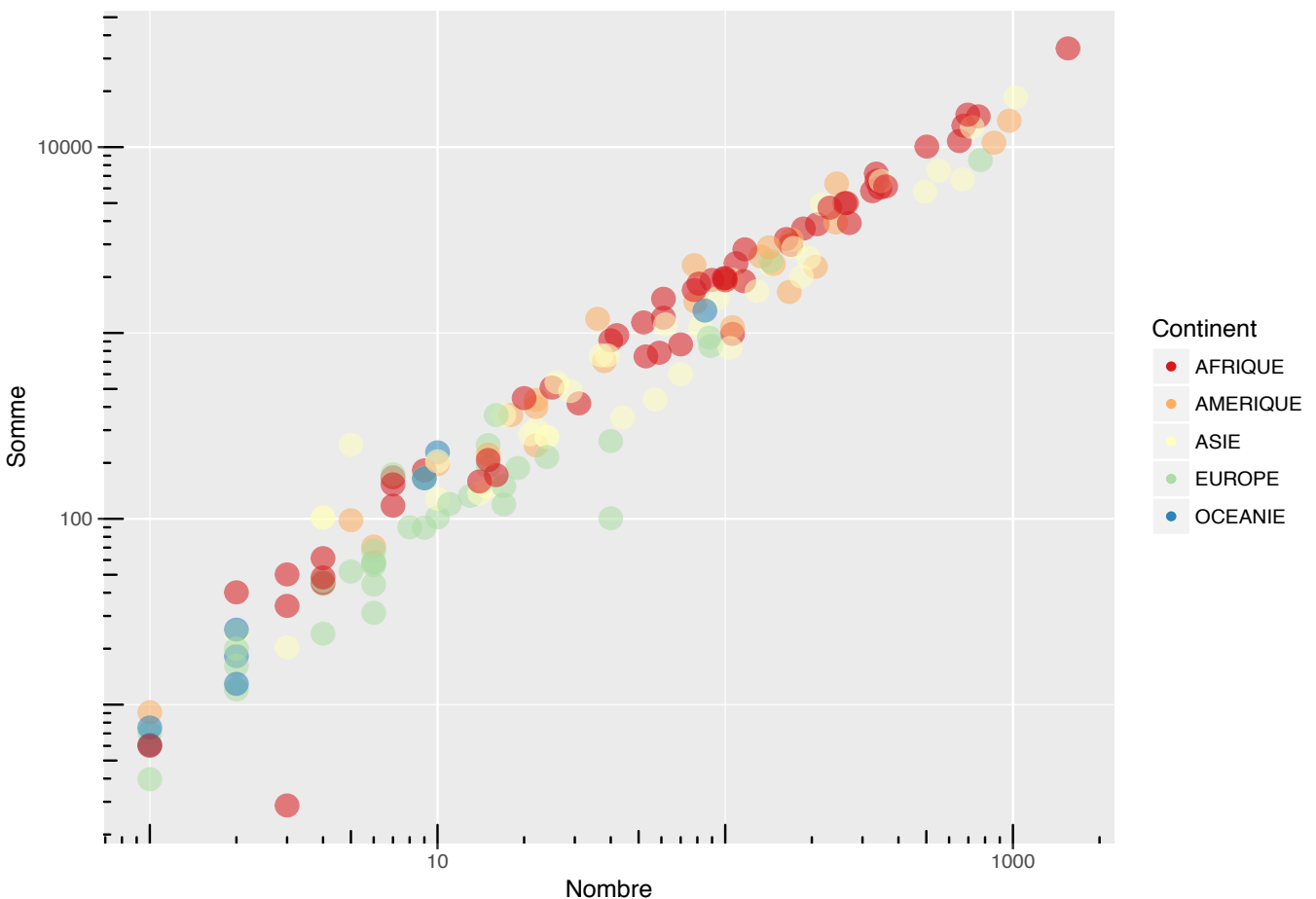
```
ggplot(durees, aes(Annee, Cumul))
+ geom_bar(stat="identity", aes(-
fill=Continent)) + scale_x_contin-
ous(breaks=c(2009:2015))
```

RELATION ENTRE NOMBRE DE MISSIONS ET TEMPS INVESTI

```
ggplot(corr, aes(Nombre, Somme)) + geom_point(aes(size=2,alpha=0.1,color=Gdes_regions)) + scale_x_log10() + scale_y_log10() +
scale_color_brewer(palette='Spectral') + guides(size=FALSE, alpha=FALSE) + labs(color="Continent") + annotation_logticks()
```

head(corr)

	Code_ISO	Gdes_regions	Region_UNU	OCDE	EEE	AD	Pays.Bas	Nombre	Somme
1	AF	ASIE	Asie meridionale	N	N	O	Afghanistan	144	2324.8333
2	ZA	AFRIQUE	Afrique australe	N	N	O	Afrique du Sud	169	2978.6333
3	AL	EUROPE	Europe meridionale	N	N	O	Albanie	13	132.9667
4	DZ	AFRIQUE	Afrique septentrionale	N	N	O	Algerie	109	2373.4333
5	DE	EUROPE	Europe occidentale	O	O	N	Allemagne	771	8516.1000
6	AO	AFRIQUE	Afrique centrale	N	N	O	Angola	61	1527.6000



La notion d'homme-mois a été problématique. Il s'agirait en fait d'un cumul de durée. Il ne faut pas la confondre avec le nombre de volontaires. J'ai cherché à savoir quelle était la relation entre le nombre de volontaires et la durée des missions. Par exemple, je me demandais si les continents qui accueillait un grand nombre de volontaires étaient aussi ceux ou ces derniers passaient le moins de temps.

Le graphique ci-dessus fait figurer tous les pays sur une échelle logarithmique. Le pays recevant le plus de volontaires est aussi celui ou la durée cumulée de toutes les missions est la plus grande. La relation est linéaire. C'est-à-dire que lorsqu'un pays reçoit un investissement en temps plus important, c'est tout simplement parce que le nombre de volontaires s'y déplaçant est plus grand. Il n'y a pas d'effet pervers inhérents à ces variables. Madagascar est suivie par le Cambodge. La Libye et la France reçoivent le moins d'investissement en temps.

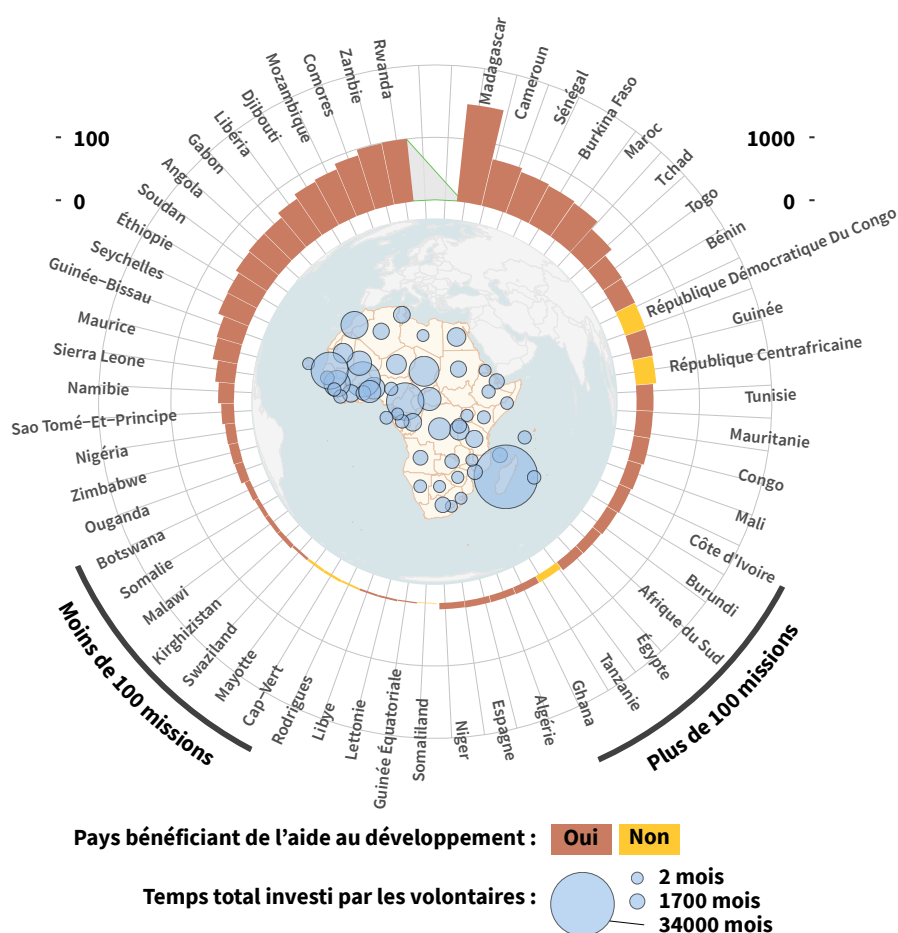
INVESTISSEMENT DES VOLONTAIRES EN AFRIQUE

```
head(corr)
  Code_ISO Gdes_regions      Region_UNU OCDE EEE AD      Pays Nombre      Somme
1      AF      ASIE      Asie meridionale  N  N  O      Afghanistan 144 2324.8333
2      ZA  AFRIQUE  Afrique australe  N  N  O      Afrique du Sud 169 2978.6333
3      AL  EUROPE  Europe meridionale  N  N  O      Albanie 13 132.9667
4      DZ  AFRIQUE  Afrique septentrionale  N  N  O      Algerie 109 2373.4333
5      DE  EUROPE  Europe occidentale  O  O  N      Allemagne 771 8516.1000
6      AO  AFRIQUE  Afrique centrale  N  N  O      Angola 61 1527.6000

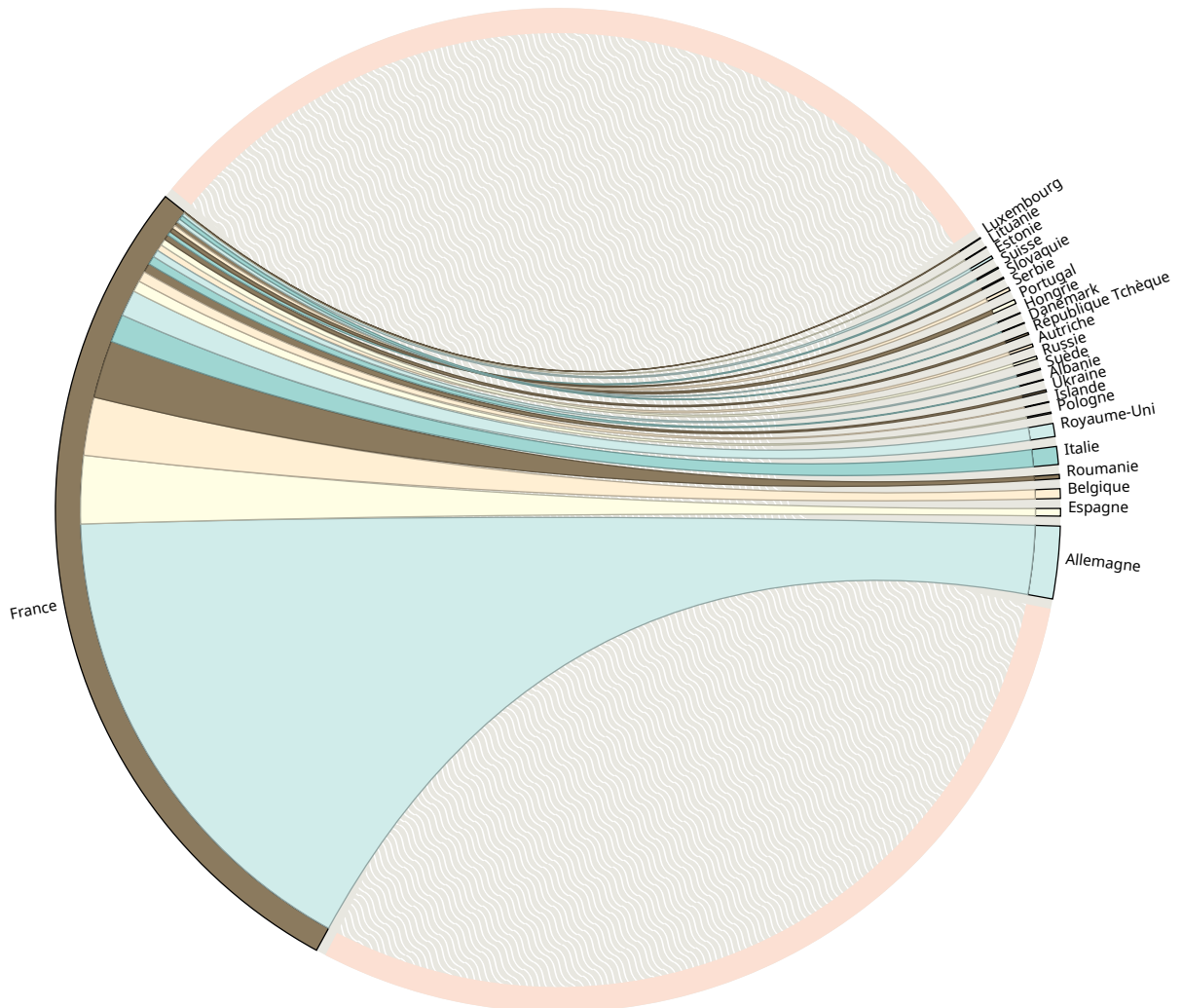
corr %>% filter(Gdes_regions=="AFRIQUE") %>% arrange(-Nombre) -> afrique
# Export pour QGIS
write.csv(afrique, "afrique.csv", row.names=F)
# Tracage des batons
afrique_inf$Nombre <- afrique_inf$Nombre*10
ggplot(afrique_sup, aes(reorder(Pays, -Nombre), Nombre)) + geom_bar(width=0.98, stat="identity", aes(fill=AD)) + coord_polar() + labs(x="", y="", fill="") + scale_y_continuous(limits=c(-3200, 1555)) + scale_fill_manual(values=ccPalette)
ggplot(afrique_inf, aes(reorder(Pays, Nombre), Nombre)) + geom_bar(width=0.98, stat="identity", aes(fill=AD)) + coord_polar() + labs(x="", y="", fill="") + scale_y_continuous(limits=c(-3200, 1555)) + scale_fill_manual(values=ccPalette)
```

J'avais pour projet de réaliser une spirale montrant le nombre d'envois dans chaque pays mais c'est une forme de représentation inadaptée au regard du nombre d'individus et de la disparité des effectifs. Je me suis rabattu sur une infographie avec un effet de superposition.

À défaut de spirale, le graphique à droite fait figurer les cinquante-six pays d'Afrique vers lesquels les volontaires sont partis entre 2009 et 2015. Comme Grégoire l'avait prédit, le résultat n'est pas excessivement intéressant. À cause de la disparité des effectifs, j'ai dû séparer les histogrammes entre les pays ayant reçu plus de cent missions et les autres. On peut voir au centre des cercles proportionnels représentant le cumul des durées des missions. La composition de l'ensemble reste à ajuster.



ECHANGES AVEC L'EUROPE DEPUIS 2011



```
# Envois depuis la France
chord_env <- read.csv("env.csv")
chord_env <- subset(chord_env, Annee>2010)
chord_env <- as.data.frame(table(chord_env$Pays))
names(chord_env) <- c("Destination", "Nombre")
chord_env$Origine <- c("France")

# Accueils
chord_acc <- read.csv("acc.csv")
chord_acc <- as.data.frame(table(chord_acc$Domicile))
names(chord_acc) <- c("Origine", "Nombre")
chord_acc$Destination <- c("France")

# Jointure
chord <- merge(chord_acc, chord_env, all=TRUE)
# Retirer les envois depuis la France vers elle-meme
chord <- subset(chord, Origine != "France" | Destination != "France")
# Reordonner les colonnes
chord <- chord[c("Origine", "Destination", "Nombre")]
# Mise en forme pour circlize
chord <- as.data.frame(table(chord$Origine, chord$Destination))
```

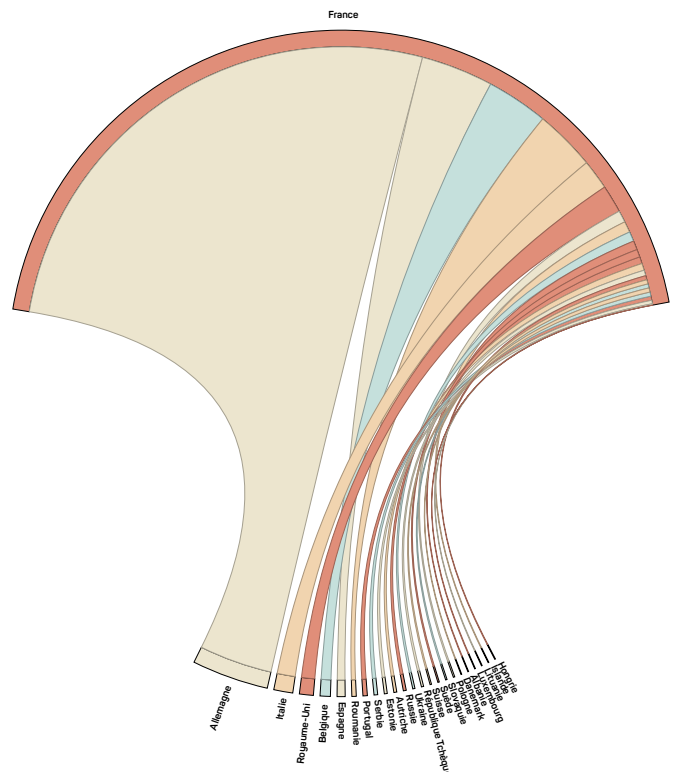
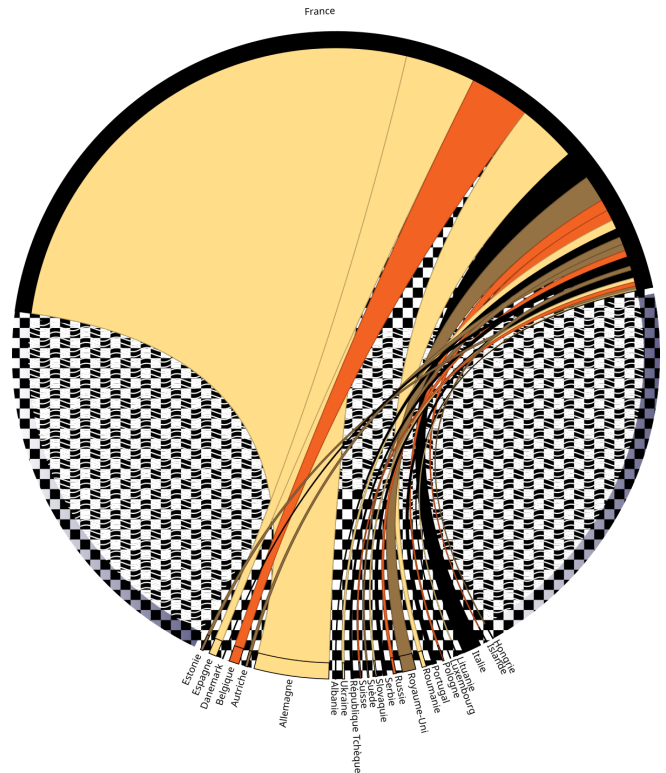
```
# Tracage avec circlize
chordDiagram(chord, transparency = 0.7)
# Export en .csv pour D3
write.csv(chord, file="chord/data.csv", row.names=F, fileEncoding="UTF-8")
```

```
Origine, Destination, Nombre
Royaume-Uni, France, 19
Italie, France, 26
Roumanie, France, 6
Belgique, France, 14
Espagne, France, 11
Allemagne, France, 105
, , 1333
France, Royaume-Uni, 40
France, Italie, 43
France, Roumanie, 88
France, Belgique, 89
France, Espagne, 104
France, Allemagne, 748
, , 1333
```

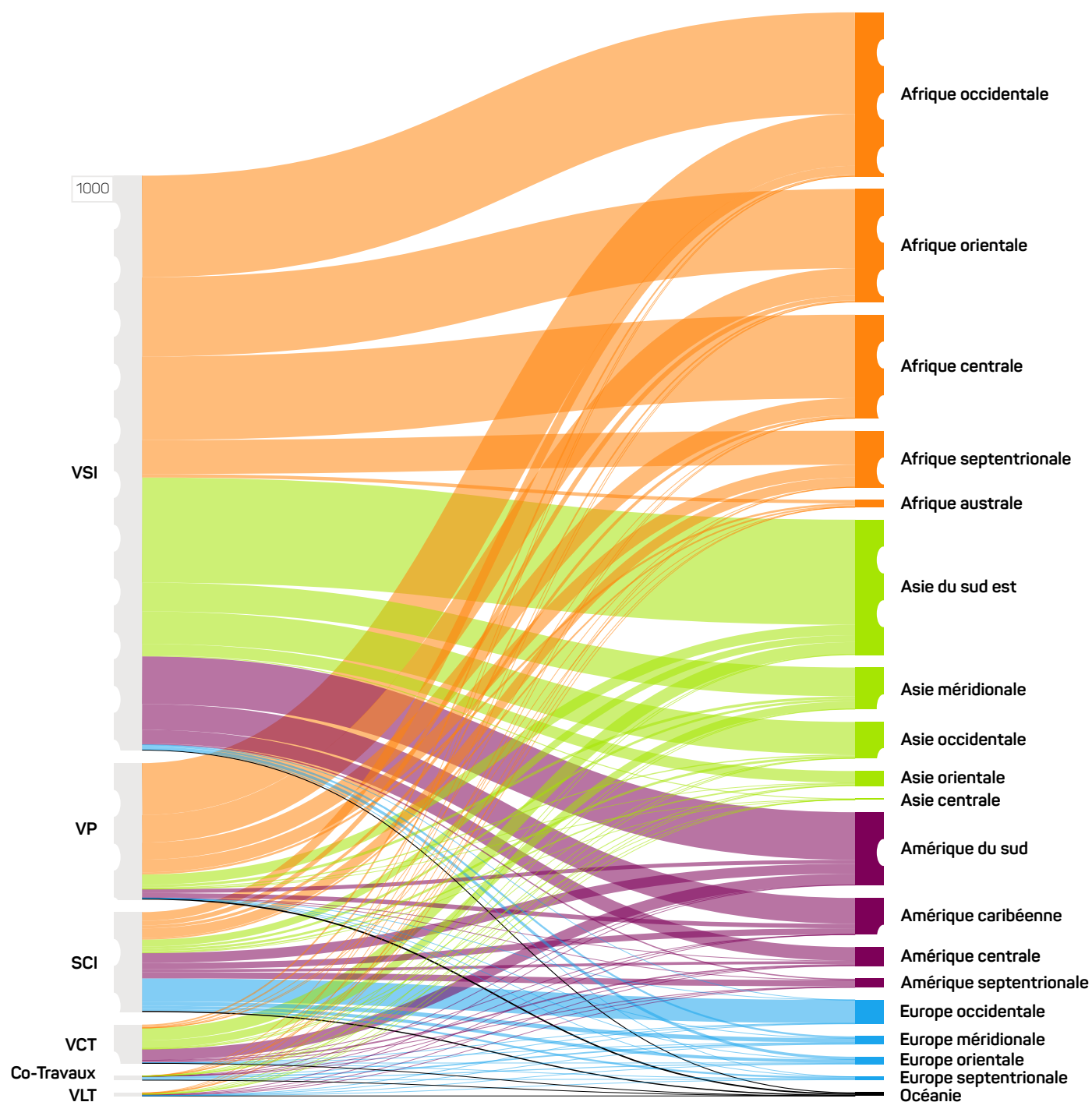
Les outils modernes de dataviz sont issus des technologies web. D3 est une librairie JavaScript qui permet de dessiner des éléments SVG dans le navigateur. En attribuant différentes variables à différents objets, il est possible de réaliser n'importe quel type de visualisation. Je ne suis pas suffisamment à l'aise en JavaScript pour créer mes propres visualisations mais j'ai pu piocher dans de nombreux exemples en ligne et adapter le code à mes besoins. Ce type de diagramme en cercle est très élégant et permet de représenter des flux entre un nombre fini d'individu. Comme d'habitude, la place disponible dicte le niveau de détail de la donnée représentée.

Ici, j'ai dû me restreindre aux pays d'Europe. J'ai inséré des individus factices pour faire prendre une forme aplatie au graphique. L'idée de départ était celle d'un arbre. Ce graphique ressemble plutôt à une étendue d'eau avec le ciel au-dessus. Tous les flux devraient être parallèles mais tout en haut il y en a six qui se croisent. J'ai tout essayé pour les réordonner, sans succès. La largeur du flux indique le nombre d'envois. On voit par exemple que la France a envoyé beaucoup plus de volontaires en Allemagne que cette dernière n'en a envoyé en France.

Une fois l'élément SVG créé, on peut copier son code dans la console du navigateur, le coller dans un fichier texte et l'enregistrer au format SVG. Il suffit de l'ouvrir dans Inkscape ou Illustrator pour modifier son esthétique. À droite, les pays du rang inférieur sont ordonnés par nombre d'envois vers la France. La forme fait davantage penser à un arbre car les individus factices sont plus petits. En revanche, la couleur n'indique aucune variable en particulier, ce qui est dommage.



NOMBRE D'ENVOIS DANS LES RÉGIONS DU MONDE

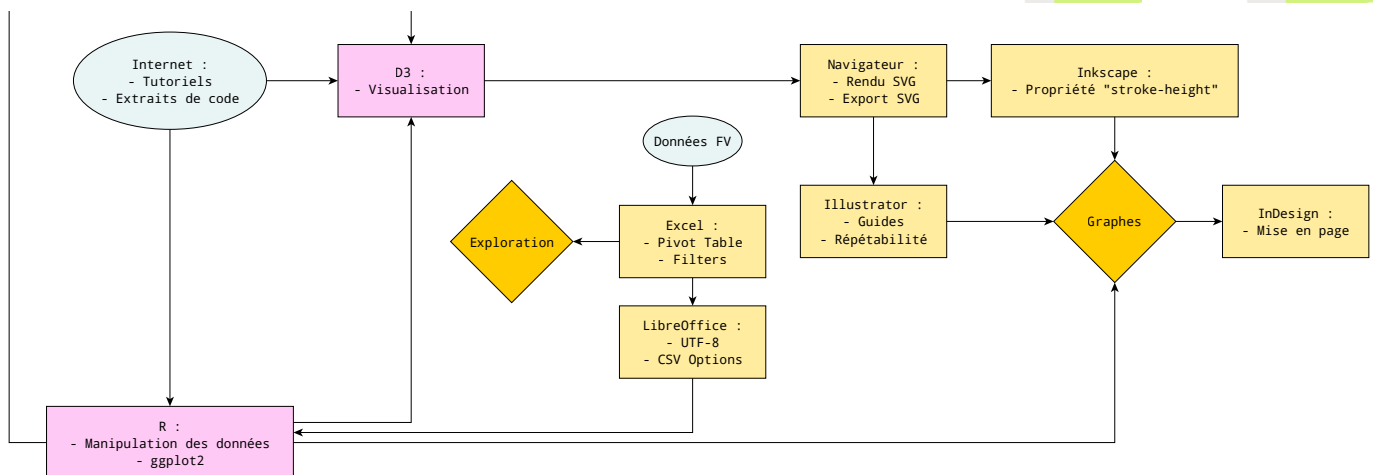
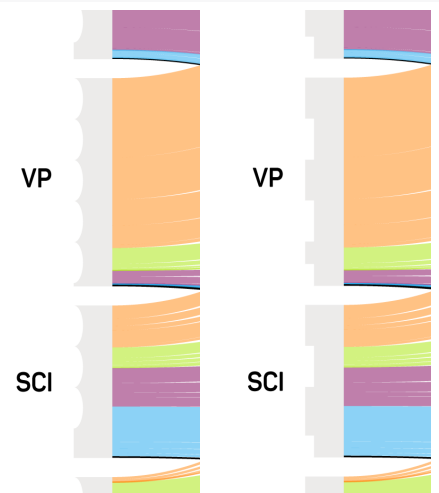



```

# Assembler les donnees
data <- count(vol, Gdes_regions, Region_ONU)
# Afficher le total pour chaque continents
vol %>% count(Gdes_regions) %>% arrange(-n)
# Ajouter un numero devant chaque continent
levels(data$Gdes_regions) <- c("1 AFRIQUE", "3 AMERIQUE", "2 ASIE", "4 EUROPE", "5 OCEANIE")
# Classer par continent puis par effectif
data %>% arrange(Gdes_regions, -n)
# Formater pour l'export vers JSON
links <- as.data.frame(env)
names(links) <- c("source", "target", "value")
nodes <- as.data.frame(c(levels(links$source), levels(links$target)))
names(nodes) <- c("names")
data <- list(links, nodes)
# Exporter
writeLines(toJSON(data, pretty = TRUE), "data.json")

```

L'idée de réaliser un graphique de ce type ("riverflow" ou "Sankey") pour l'atlas est venue dès le premier jour. Après des essais infructueux avec la librairie googleVis, j'ai réussi à adapter un morceau de code produit par Mike Bostock, créateur de D3. Suite aux remarques d'Olivier Pissot, j'ai réordonné les "nodes" de part et d'autre afin qu'elles apparaissent dans un ordre décroissant. J'ai ajouté manuellement dans Inkscape des graduations pour donner un ordre de grandeur. Illustrator ne gère pas la propriété "stroke-height" du standard SVG et donc ne peut pas faire apparaître les flux.



Au final, je suis très content d'avoir pu me familiariser avec ces nouveaux outils. Dans l'avenir j'aimerais mieux les maîtriser et apprendre à travailler plus vite. Je fais beaucoup d'aller-retours entre Excel/Calc et R car je ne suis pas à l'aise sur les opérations de "merging". Parmi les outils m'ayant bien servi au cours du semestre il y a également Notepad++ (avec TextFX), LyX et InDesign.

BROUILLONS

The dashboard displays the following visualizations:

- Sankey Diagram:** Shows the flow of data between various countries, including France, Hongrie, Islande, Italie, Lituanie, Luxembourg, Pologne, Portugal, Roumanie, Royaume-Uni, Russie, Serbie, Slovaquie, Suède, Suisse, République Tchèque, Bosnie-Herzégovine, Bulgarie, Croatie, Finlande, Grèce, Irlande, Kosovo, Lettonie, Malte, Monténégro, Norvège, Pays-Bas, Slovaquie, Yougoslavie, Albanie, Allemagne, Autriche, Belgique, Danemark, Espagne, Estonie, and France.
- Map of Morocco:** Shows the geographical distribution of data points across the country, with labels for various cities and regions.
- Ternary Plot:** A triangular plot showing the distribution of data points across three axes: Age, Duree, and Année. The plot is divided into regions: AFRIQUE, AMERIQUE, ASIE, EUROPE, and OCEANIE.
- Bar Chart:** Shows the distribution of mission duration (Duree_mission_mois) over time (Année) for various regions: AFRIQUE, AMERIQUE, ASIE, EUROPE, and GLOBAL.
- Sankey Diagram:** Shows the distribution of VSI (Valeur de Satisfaction Individuelle) across different regions: AFRIQUE, AMERIQUE, ASIE, EUROPE, and OCEANIE.
- Chord Diagram:** Shows the distribution of VSI across different regions, with labels for various countries and regions.
- Bar Chart:** Shows the distribution of VSI across different regions, with labels for various countries and regions.