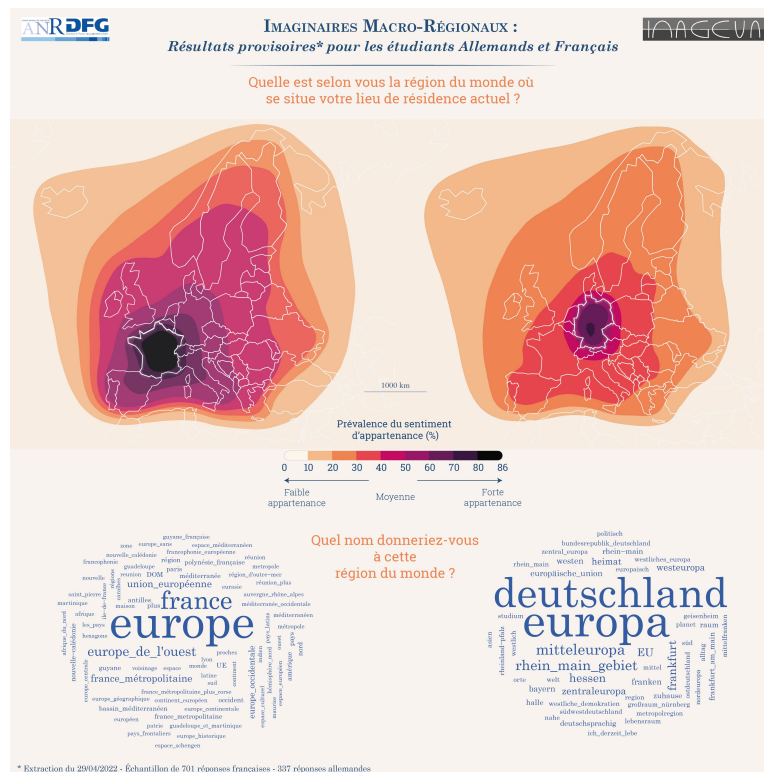


Gestion d'une base de données qualitative et géométrique



Elina Marveaux

Collège International des Sciences Territoriales
Encadré par Hugues Pécout - Ingénieur de recherche, géomaticien

Master Carthagéo

12 Juin 2022

Résumé du stage (dont le sujet) en 10 lignes en français et en anglais. Informatif et concis, ce résumé doit refléter l'esprit du document, définir les buts et les méthodes, les résultats et les conclusions. Il se présente sous la forme d'un paragraphe unique, sans alinéa.

Table des matières

Introduction	1
Notes rendez-vous Antonine	3
Les visualisations	3
Méthodologie	3
Analyse du besoin en amont ?	4
Livrables	4
GIT	4
Gestion de projet / Workshop	5
La forme du mémoire	5
1 Contexte	7
1.1 EuroBroadMap	7
2 Méthodologie	9
2.1 La chaîne de traitement	9
2.2 Traitement de la base de donnée “géométries”	9
2.2.1 Géométries et leur classification : enjeux méthodologiques	9
2.2.2 La visualisation pour exploration	10
2.2.3 Garder ou non les polygones	11
2.3 L’outil : Maptionnaire	12
2.3.1 Les biais	12
2.3.2 Les questions	13
2.4 Exploitation de la Base de donnée “Réponses”	13
3 Résultats	15
3.1 Poursuites et valorisation	15
3.2 Data	20
Conclusion	21
Remerciements	23
References	25
Appendices	25
Glossaire	27
Quelques résultats supplémentaires	29

Introduction

Elle doit mentionner : les objectifs, les lieux d'étude, l'intérêt du projet : pour un public particulier ? Une nouvelle méthodologie ? Une monographie sur un espace donné ? Durée. Un rapport orienté « recherche » développera la problématique qui sous-tend le questionnement de recherche et formulera, le cas échéant, les hypothèses de travail.

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 1 + 1

[1] 2

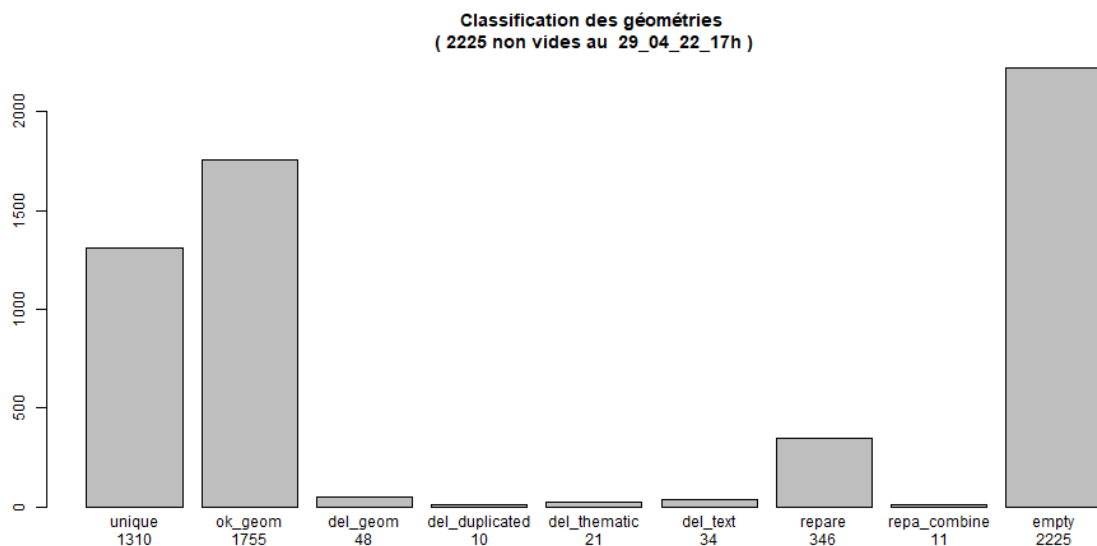


Figure 1: Chart Bar des NAS

Notes rendez-vous Antonine

Dimension politique des visualisation à mettre en avant

Les visualisations

Faire la différences entre les visualisations des polygone internes pour leur diagnostic et la présentation des “problèmes” et la représentation finale, les cartographies qui doivent être finies et présentées comme résultat et non pas comme exploration.

Les deux modes de représentations font appel à des enjeux différents (ainsi que des publiques, objectifs)

Méthodologie

- Décrire la méthodologie
- Quelle est la qualité de l'information :
- Qualité du point de vue de la géomatique
- Quels biais introduits dans les enquêtes “habituelles” ou en tout cas celle précédente celle-ci (eurobroadmap)
- Quel biais introduits par la méthodologie actuelle, quel enjeu vis à vis du choix de l'outil
- Comment sont identifiés/détectés les problèmes dans la BD et BDG, quelles propositions sont avancées, lesquelles sont mises en place et comment trancher ? **proposer une chaîne de traitement sous forme d'image”
- Voir Lena Sanders : modèle LOGIT pour explorer la significativité ou non du univ_field

- Quelle pondération de polygones lorsqu'ils se superposent, lorsqu'ils n'ont pas la même taille mais décrivent des ensembles similaires (pays) ou lorsqu'il sont multi-polygones donc séparés et parfois se recouvrent
- Comment analyser les aires disjointes ou emboîtées ?

Analyse du besoin en amont ?

Livrables

- Quels sont les produits à réalisés et à qui s'adressent-ils ?
- Priorité sur les **géométries** et leur enjeux :
- plusieurs dessins sont autorisées
- plusieurs modes de saisi sont possibles chacun avec leur biais
- Plusieurs problèmes en découlent
- Base de donnée Géo et Respondent avec documents de présentation
- explorations de la BDG :
- Visualisations exploratoires et travail de représentation pour le diagnostic (explorer les chorèmes pour proposer une typologie des polygones et relations rencontrées dans la BDG)
- Visualisations finales / exploratoires en tant de résultats de l'exploration : on parle de représentations visant à proposer certaines interprétations des résultats (sous formes de prototypes car la valorisation finale est attendue en dehors du cadre du stage)

GIT

- expliquer l'intérêt de l'outil
- Faire la différence entre 'usage généraliste de l'outil et l'usage qu'on en fait en interne (dans le care du stage notamment)
- Expliquer pourquoi une page web et quel est l'intérêt (éventuellement pour mettre à disposition des résultats du stage et la présentation du mémoire ar exemple)

Gestion de projet / Workshop

Ici discuter des apports du stage dans le domaine de la gestion de projet et de la mise en place d'un workshop. - Dans quelle mesure suis-je force de proposition - Quel suivi, même informel est mis ou ai-je mis en place ? Finalement le suivi est assez souple, je suis libre de faire ce que je veux mais le cadre est bien posé et les objectifs déterminés. Par exemple, j'ai été accompagnée dans la découverte de la BD e des premiers scripts exploratoires, j'ai eu comme objectif de réaliser un document de présentation de cette base de donnée et des traitements qui y ont été réalisés pour la rendre propre ou qui peuvent être réalisés dessus en guise d'exploration. J'ai le champ libre pour la réalisation de ce document concernant à la fois les outils que j'utilise et ce que j'y met. Je reste encadrée lorsque j'ai des questions et ait besoin d'être recadrée.

- Mettre ne avant l'informel : les formations et les autres projets/ présentations auxquelles j'ai été invitées à participer
- Parler des compétences acquises en matière de logiciels/ technique (quarto, rédaction latex/yaml, regex, git)

La forme du mémoire

- utiliser souvent les illustrations et les graphiques, notamment pour les chaines de traitement et l'organisation. il est important que le mémoire reste lisible et très rapidement compréhensible.
- **Etat de l'art** : Surtout un état de l'existant notamment en matière d'outils et de méthodologie. Le stage reste un stage ingénieur, et non pas de recherche. Il faut mettre en avant ce qui a déjà été produit en terme technique (exemple : si je rencontre un problème, la solution existe probablement déjà, il faut alors que je fasse état de cette ou ces solutions et que je discute de ce que j'en fait, comment je l'adapte à mon travail etc
- Deux options pour le plan :
- Classique : 3 parties reprenant chacune tous les objectifs / sous projets

- Contexte (structure, projet, enjeux du projet, comment le stage s’insère dans le projet, quels acteurs sont rencontrés et quels sont leurs rôles)
 - Méthodologie mise en place
 - Résultats avec mise en perspective, retour sur le travail effectué
- ou alors Une partie par projet / Objectif reprenant chacune les parties précédemment décrites
 - 1 : exploration BDR : Contexte/méthodologie/résultats
 - 2 : Exploration des géométries Contexte/méthodologie/résultats
- Bien mettre en avant ce que j’ai découvert, ce que j’ai **proposé** les difficultés rencontrées et les problèmes soulevés même s’ils n’ont pas été résolus
- **OBJECTIFS** : reproductibilité, analyse, mise à disposition des données (et proposition d’exploitation de ces données)

1 Contexte

Contexte (structure, projet, enjeux du projet, comment le stage s'insère dans le projet, quels acteurs sont rencontrés et quels sont leurs rôles)

1.1 EuroBroadMap

Les résultats Europe sont très normées, voir prévisibles, tandis que d'autres pays adoptent un regard plus critique vis à vis de l'Europe. - Chine mode, luxe, sp Afrique : Tunisie et Sénégal : critique sur le racisme / inégalité / colonialisme Brésil :

2 Méthodologie

Contenu central du rapport

Il comporte deux ou trois parties (1, 2, 3) subdivisées en deux ou trois sous parties (1.1, 1.2, 1.3) tout au plus. En lisant les titres des 4 à 9 sous parties, on doit pouvoir se faire une idée du contenu et de la démarche de l'étude. Le texte est fait pour être lu par une personne qui n'a pas suivi le travail du stagiaire. Un rapport orienté « recherche » intégrera obligatoirement une partie « état de l'art » et valorisera les références bibliographiques.

2.1 La chaine de traitement

2.2 Traitement de la base de donnée “géométries”

Possibilité de créer un package pour exploiter les données.

Pour produire des tableaux à la volée (BDG & BDR) et les graphiques, cartes, statistiques qui en découlent. Ce travail nécessite l'identification précise de productions pour chacun de ces volets (quels graphiques sur quelles données, etc).

2.2.1 Géométries et leur classification : enjeux méthodologiques

- Comparer la superficie du polygone avec la superficie réelle de l'espace décrit (notamment lorsque le mot attribué correspond à cet espace).
- Quel ordre des polygones pour un meme mot donnée
- Justifier l'interet de juger à l'oeil (controle humain face à la redondance de la tache).

2 Méthodologie

- Justifier l'usage des menus et leur pertinences vis à vis d'une approche automatisée (identifier les actions automatisables)
 - Geom repair : intersection
 - txt move/add : au cas ou mais tres peu utilisé finalement
 - Typologie : difficile de faire une typologie automatisée (on peut imaginer un scale comme etant une hierarchie des superficie lorsqu'elles sont contenues les unes dans les autres à x%, mais le multi unique et le overlap sont plus complexes car relevent d'une appréciation de l'intention du dessinateur (l'intersection partielle de deux ensemble peut etre due à une imprecision comme à une volonté du dessinateur))

Il reste de nombreuses actons répétitives voire inutiles dans la boucles.

On peut envisager une comparaison des résultats automatisés et manuels

Classification manuelle supervisée (à deux ou trois) pourquoi ? quelle utilité, comment s'organise-t-on pour controler les réponses et les standardiser ?

2.2.2 La visualisation pour exploration

Décrire la méthode d'affichage et de classification des polygones puis exposer les problèmes Présenter les problèmes de géométrie sous forme de schémas / chorème ?

Problème sémiologiques pour juger rapidement et efficacement dans un premier temps du polygone, puis dans un second temps des choix fait sur ce/ces polygones (contrôle).

Le choix du fond de carte et de sa généralisation posent des enjeux concernant d'une part les polygones décrivant des territoires insulaires qui n'apparaissent pas et d'autre part des polygones décrivant des petits territoires type villes dont on ne peut distinguer les contours.

Dans le premier cas, on visualise un polygone dans un océan sans repères autour (pas d'îles), on peut estimer la justesse du dessin grace aux informations récupérées directement du répondant (le mot associé ou son université d'appartenance surtout) ainsi que par les autres polygones dessinée (lorsqu'il y en a). Surtout ce problème est corrigé en choisissant

une échelle de représentation plus fine pour le fond de carte.

Dans le second cas le problème est plus délicat. Il ne s’agit pas d’un niveau de détail du à la généralisation mais à des limites administratives ou de toponymes présents ou non pour aider à reconnaître le lieu décrit par le polygone. Dans ce cas, choisir un fond de carte plus riche pose d’abord la contrainte de l’alourdissement de la représentation, de la saturation de l’espace visuel au détriment du jugement du polygone (qui est l’objet premier nécessitant toute l’attention du visualisateur). Cette option améliore la reconnaissance du lieu décrit par le polygone mais pas à tous les coups. Il peut encore être nécessaire, lorsque la région n’est pas connue, ou encore trop petite de “dézoomer”.

L’alternative retenue (dans un premier temps) a été de quitter la boucle d’affichage et de passer par une cartographie interactive des tous les polygones d’un répondant sur une même carte. Ici le fond de carte est produit par Open Street Map et la généralisation des toponymes et des limites administratives des pays se fait automatique selon le zoom répondant de façon interactive. Ici encore les informations additionnelles données par le répondant sont accessibles pour chacun des polygones via une infobulle générée au clic.

Une autre alternative eu été de directement représenter les polygones via cette carte interactive au sein même de la boucle.

Cette option n’a pas encore été explorée.

2.2.3 Garder ou non les polygones

- Quel choix pour les **polygones trop petits** : au dessus du quartier ; problème lors de la représentation (un carreau entier selon la résolution pour une surface potentiellement plus petite que ce carreau) mais ne peut pas être le seul argument... considère-t-on que quelqu’un entourant “sa maison” répond bien à la question “qu’elle est la zone d’appartenance de votre pays ?”. On peut se demander si la façon dont est posée la question induit cette réponse et donc si elle doit être disqualifiée ou non. un polygone trop petit pose aussi la question de l’anonymat et de la protection du répondant.
- Problème pour les polygones multi-unique et scale, ainsi que over et scale (lorsque les polygones répondent à plusieurs logiques d’appartenance)

Table 2.1: Systèmes d'opération compatibles au 28/06/2022

(a) Compatibilité ordinateur		(b) Compatibilité smartphone	
WINDOWS	MACOS	Android	iOS/iPAD
Windows 10 : 2014	macOS 11 : 2020	Android 11 : 2020	iOS/iPadOS 14 : 2020
Windows 11 : 2021	macOS 12 : 2021	Android 12 : 2022	iOS/iPadOS 15 : 2021

- Pour les polygones “world” et leur probleme de plot peut-on tous les remplacer par un polygone type ? (plutot non)
- Que faire des polygones dupliqués

2.3 L'outil : Maptionnaire

2.3.1 Les biais

2.3.1.1 Systèmes d'exploitation et navigateurs internet pris en charge

Does Maptionnaire have any system or browser requirements? *Maptionnaire uses commercially reasonable efforts to support the two most recent major versions of operating systems with significant market share running up-to-date versions of browsers with significant market share*

As of November 9, 2021, the supported operating systems and versions for respondents are: - Windows (11, 10) - macOS (12, 11) - Android (12, 11) - iOS/iPadOS (15, 14)

The supported browsers are Chrome, Safari, Firefox, Samsung Internet, Edge, and Opera. For optimal performance, please remember to make sure that you have the latest browser version.

If you are part of a team using Maptionnaire to create questionnaires and other content, we recommend that you use the latest version of Chrome.

Les tables suivantes décrivent l'année de sortie des OS pour les ordinateurs et smartphone.

Comme on peut le voir sur la Table 2.1, on peut imaginer des problèmes de compatibilité peut être avec certains parcs informatiques/technologiques possiblement contrôlable avec des données économiques et de part de marché pour chaque pays. L'application a été massivement (*chiffre étude*)(Bailly 2005) utilisée au moyen d'un téléphone, la Table 2.1b nous montre que pour bien fonctionner il faut des téléphones plutôt récents, de moins de 2 ans... Ou en tout cas fonctionnant sur un OS récent (de nombreux anciens téléphones peuvent soutenir de telles mises à jour mais pas tous).¹

Test de l'application évisageable sur des sites de développement web type “browser-Stack”.

2.3.2 Les questions

DROM On peut supposer que la surreprésentation des droms dans les polygones des français métropolitains est due à la récurrence du terme “DROM” tout au long du questionnaire

Polygones La question n'est pas exactement traduite correctement dans chaque langue. Des mots ont été ajoutés en allemand par exemple. D'autre part les termes “pays” et “limites” posent des problèmes d'interprétation (pays = country = campagne ??? - limite = frontière ???)

2.4 Exploitation de la Base de donnée “Réponses”

- Explo des données manquantes : VIM ?
 - distribution des VM (dispositifs et mécanismes)
 - CAH ou autre pour établir une typologie des NA (profils des répondants)
- Objectifs de l'enquête en termes d'effectifs :
 - 5 pays : Tunisie, Turquie, France, Allemagne, Irlande
 - 3 Villes par Pays
 - 240 étudiants par ville
 - 40 étudiants par discipline

¹Ainsi les téléphones Samsung et iPhone peuvent être plus performants et mieux mis à jour que des marques moins importantes n'assurant pas la compatibilité au cours du temps. (Noucher 2015)

3 Résultats

Ils comportent les éléments qui permettent d’apprécier si la démarche, la méthode, etc... sont utilisables, généralisables

3.1 Poursuites et valorisation

On peut envisager la création d’un package R permettant l’exploitation de cette base de donnée. Il reprendrait l’essentiel des scripts déjà écrit pour la construction de la BD et proposerait d’autres fonctions permettant d’une part l’extraction de sous bases de données (spécifiques à des ensembles géographiques ou à certaines thématiques issues des questions) et d’autres part une exploitation visuelle sous forme de graphiques et de cartes “clé en main”.

L’intérêt d’un package :

- pas dans une culture de remise à disposition avant mais maintenant oui
- Peu d’autres projets qui reposent sur les projets initiaux
- ANR corpus : produit central est la donnée et le but est justement d’être utilisé par tous
- Aussi pu réutiliser à l’intérieur du projet : problème de temporalité et des objectifs directs.
- meta analyse sur l’usage global de toutes les données produites ?????

Faire la différence entre valorisation/diffusion et “reproductibilité”

This is a Quarto book.

3 Résultats

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 1 + 1

[1] 2

Ici la Figure 3.1 produit avec R puis enregistré en image et lu en markdown

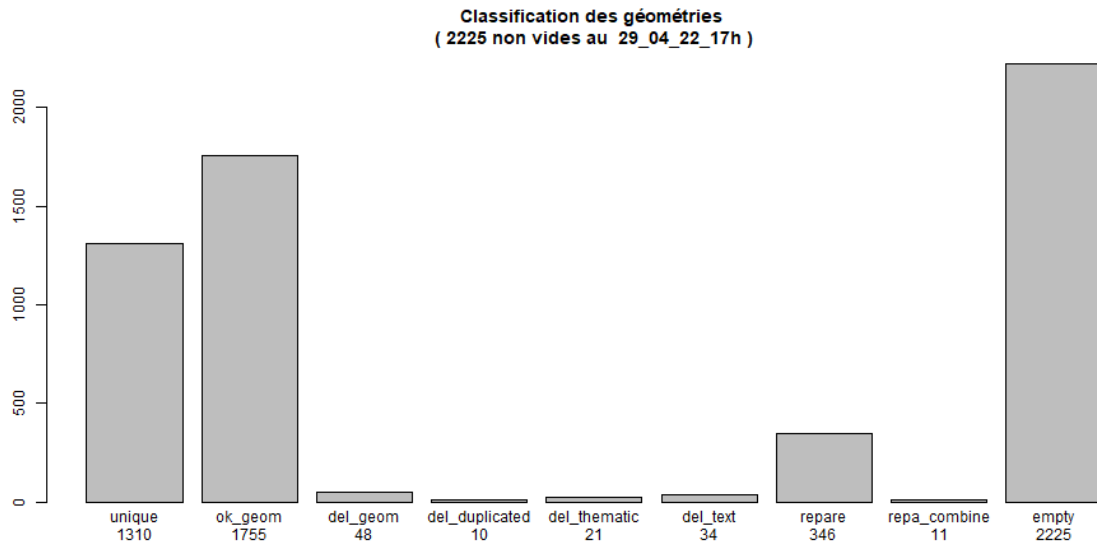


Figure 3.1: Chart Bar des NAS

Puis ci-dessous la Figure 3.2 image lu avec `knitr` via un chunk r

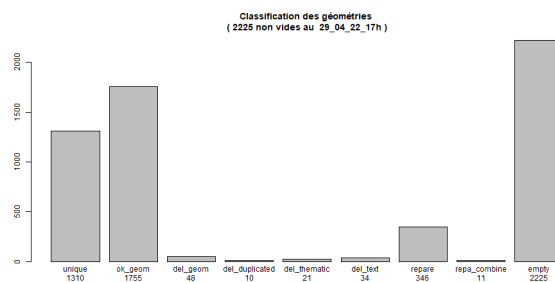
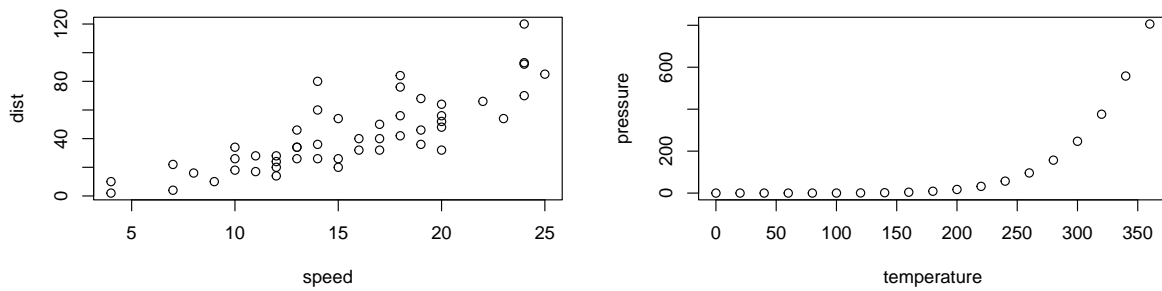


Figure 3.2: Chart with Knitr

Voici le Figure 3.3 produit avec R et dont on ne peut pas lire le (court) script et qui s'affiche sur deux colonnes

Enfin la Figure 3.4 décrivant une courbe et son loess ainsi que le script le construisant



(a) Speed and Stopping Distances of Cars (b) Vapor Pressure of Mercury as a Function of Temperature

Figure 3.3: Two sided charts

```

1 mtcars_jj <- mtcars
2
3 mtcars_jj$am <- factor(x = mtcars$am,
4                       labels = c("automatic", "manuel"))
5
6 ggplot(data = mtcars_jj, aes(hp, mpg, color = am)) +
7   geom_point() +
8   geom_smooth(method = "loess") +
9   theme(legend.position = "bottom")

```

3 Résultats

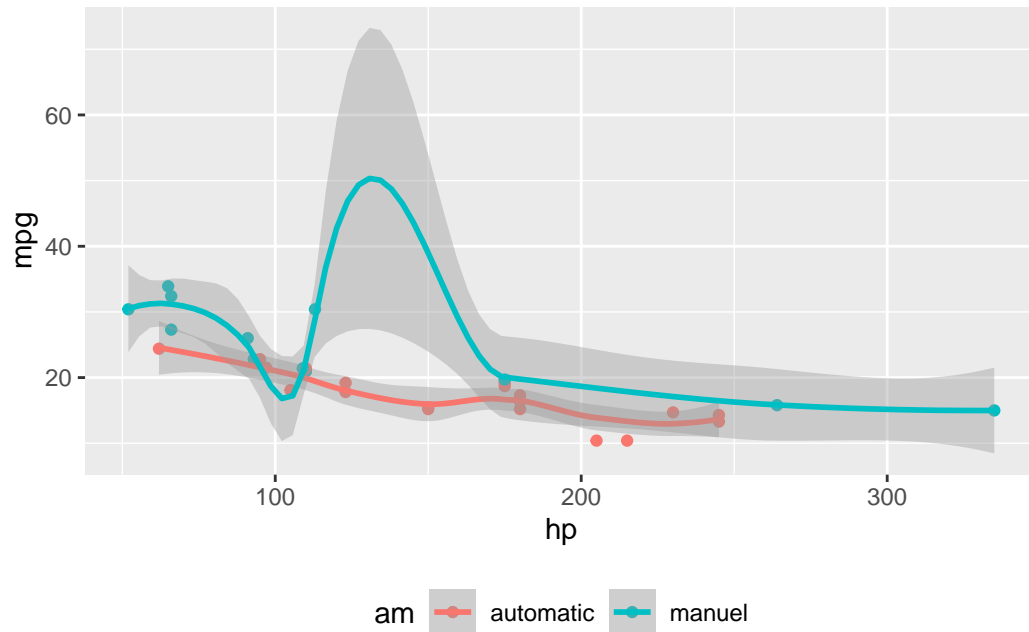
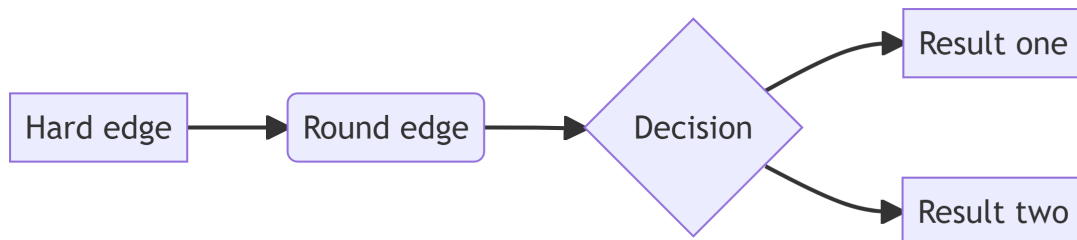


Figure 3.4: MPG gorsepower

Ceci est un exemple de diagramme écrit en mermaid

```
1
2 flowchart LR
3   A[Hard edge] --> B(Round edge)
4   B --> C{Decision}
5   C --> D[Result one]
6   C --> E[Result two]
```



Un autre exemple de diagramme

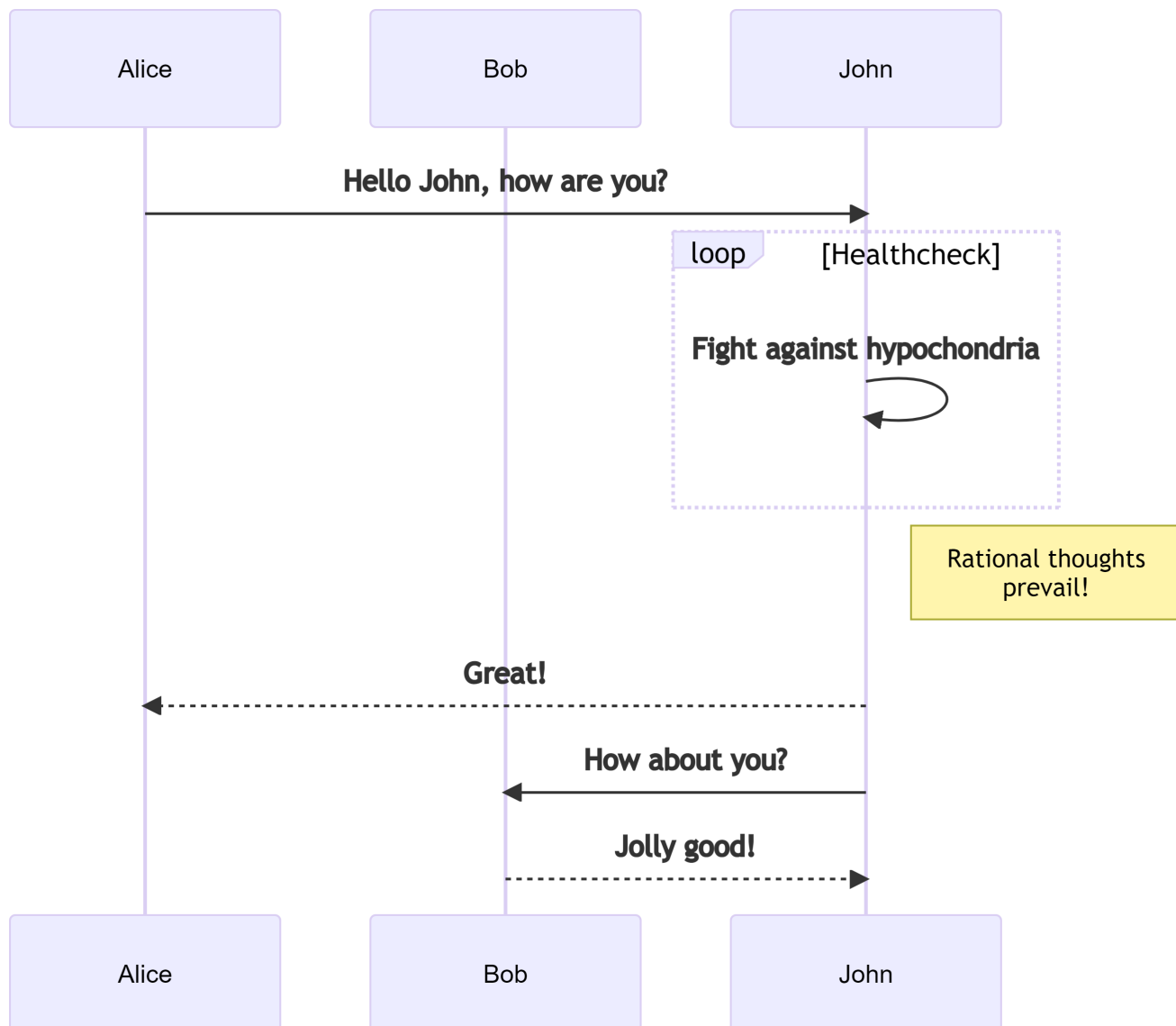
```
1
2 sequenceDiagram
3   participant Alice
4   participant Bob
```



```

5  Alice->>John: Hello John, how are you?
6  loop Healthcheck
7    John->>John: Fight against hypochondria
8  end
9  Note right of John: Rational thoughts <br/>prevail!
10 John-->>Alice: Great!
11 John->>Bob: How about you?
12 Bob-->>John: Jolly good!

```



mermaid diagram

3.2 Data

The penguins data from the **palmerpenguins** package contains size measurements for `nrow(penguins)` penguins from three species observed on three islands in the Palmer Archipelago, Antarctica.

The Figure 3.5 below shows the relationship between flipper and bill lengths of these penguins.

```

1 ggplot(penguins,
2       aes(x = flipper_length_mm, y = bill_length_mm)) +
3   geom_point(aes(color = species, shape = species)) +
4   scale_color_manual(values = c("darkorange", "purple", "cyan4")) +
5   labs(
6     title = "Flipper and bill length",
7     subtitle = "Dimensions for penguins at Palmer Station LTER",
8     x = "Flipper length (mm)", y = "Bill length (mm)",
9     color = "Penguin species", shape = "Penguin species"
10  ) +
11  theme_minimal()

```

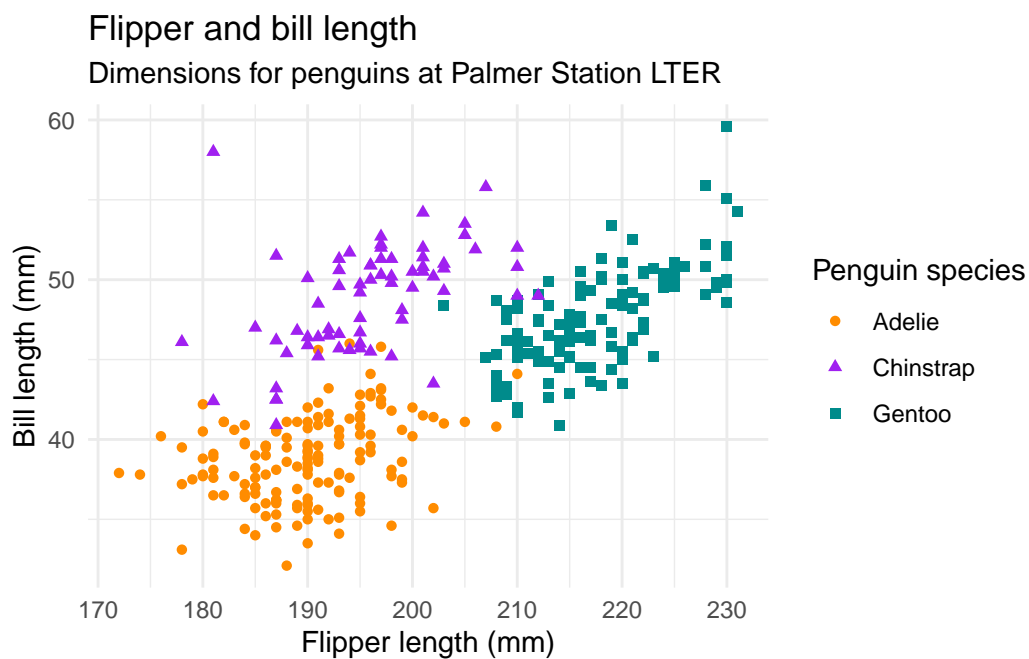


Figure 3.5: Happy Penguins

Conclusion

Dans quelles conditions peut-on la généraliser ? Quelles réponses ont été données ? Quelles sont celles qui ne l'ont pas été ? Que faut-il envisager pour l'avenir ?

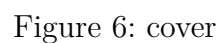
Remerciements

References

- Bailly, Antoine. 2005. "Voyage en Géographie." *BSGLg*, January. <https://popups.uliege.be/0770-7576/index.php?id=1857>.
- Noucher, Matthieu. 2015. *De la trace à la carte et de la carte à la trace : pour une approche critique des nouvelles sources de fabrique cartographique*. Presses des Mines. <https://halshs.archives-ouvertes.fr/halshs-01212022>.

Glossaire

Some results that wouldn't fit into the main thesis



Gestion d'une base de données qualitative et géométrique

Elina Marveaux

12 Juin 2022

Résumé du stage (dont le sujet) en 10 lignes en français et en anglais. Informatif et concis, ce résumé doit refléter l'esprit du document, définir les buts et les méthodes, les résultats et les conclusions. Il se présente sous la forme d'un paragraphe unique, sans alinéa.

MOT CLEF

- paradonnées
- metadonnées
- reproductibilité
- base de donnée
- géomatique
- programmation lettrée