

# Analyse des données STRVA, une aide dans la compréhension de l'utilisation du territoire.

AUTEUR : PAULINE DE MONTLEBERT

DATE :19/10/20

MOTS CLÉS : CARTOGRAPHIE, AMÉNAGEMENT, ANALYSE, SIG, GPS

De nombreuses personnes partagent des informations géolocalisées à travers des sites communautaires dédiés, comme l'application STRAVA. Il s'agit d'une application permettant aux coureurs de déposer leurs traces GPS afin de suivre leurs performances. Ces données peuvent être riches d'informations sur l'usage du territoire et sur le comportement des usagers.

# Introduction

---

## Contexte

De nombreuses personnes partagent des informations géolocalisées à travers des sites communautaires dédiés. C'est par exemple le cas des coureurs à pied, qui partagent les traces GPS de leurs courses à travers le site de Strava.

Ces données peuvent être riches d'informations sur l'usage du territoire et le comportement des coureurs. Elles peuvent contenir des informations sur les habitudes des coureurs et peuvent être utilisées dans le but d'aménager les accès aux zones de courses.

## Objectif

Le but de ce projet est d'analyser les données STRAVA pour comprendre le comportement des coureurs sur les zones définies avec le commanditaire : le parc de Vincennes. Cela permet ainsi d'effectuer des cartes montrant par exemple les points de départ et d'arrivée des coureurs, le sens de circulation et d'autres analyses qui pourront à l'avenir servir lors de l'aménagement des territoires et ainsi faciliter l'accessibilité des principales zones de courses empruntées par les coureurs.

# Travail réalisé

---

## Approche suivie

La première étape du projet consistait à récupérer les données STRAVA sur l'île de France. Il a fallu ensuite filtrer ces données pour ne récupérer que les traces GPS couvrant le parc de Vincennes et le lac de Vaires-sur-Marne.

Après le nettoyage des données, l'analyse des traces GPS a pu commencer. Pour cela plusieurs cartes ont été réalisées pour comprendre le comportement des coureurs :

- une première carte sur les points de départ et d'arrivée, une deuxième sur le sens de circulation des coureurs, et une carte sur les périodes de courses (semaine ou week-end).

Pour la réalisation des différentes cartes, plusieurs logiciels ont été utilisés :

- Le filtrage des données a été réalisé sous FME qui est un logiciel qui permet une synchronisation massive d'informations d'une source de données vers une autre. Il était nécessaire d'utiliser FME pour le filtrage des données. En effet, contenu du très grand nombre de données couvrant toute l'île de France, un simple logiciel de SIG tel que QGIS ou Arcgis-pro n'aurait pas pu réaliser le filtrage correctement.
- Le choix du logiciel s'est porté sur QGIS car c'est un logiciel gratuit et plus intuitif pour les traitements que Arcgis-pro.

Le travail s'est organisé autour de plusieurs questions :

1. Quelle est la zone d'étude définie ? (Filtrage des données de Vincennes et Torcy)
2. D'où partent et arrivent les coureurs ? (Station de métro, habitation, ou autre)
3. Existe-t-il un sens privilégié par les coureurs ?
4. Où sont pratiquées les courses longues et les courses courtes ?
5. Comment sont réparties les courses durant la semaine (week-end ou semaine) ?

## Réalisation

Durant ce projet toutes les données shapefiles utilisées sont définies en Lambert 93 et orientées au Nord.

### *Quelle est la zone d'étude définie ?*

Dans un premier temps il fallait créer un polygone assez large autour du bois de Vincennes

Une fois la zone d'étude définies approximativement il fallait les affiner en ne prenant que le contour du parc. Ce traitement a été effectué sur QGIS.

### *D'où partent et arrivent les coureurs ?*

Une fois les traces GPS de la zone extraite, les premières cartes d'analyse on put être réalisées. Les premières montrent les points de départ et d'arrivée de chaque trace GPS. Ce traitement a été effectué sur QGIS à l'aide de la calculatrice de champ et des expressions «  $xat(0)$  » et «  $yat(0)$  » qui nous renvoient les coordonnées du premier point d'une polyligne. A l'inverse «  $xat(-1)$  » et «  $yat(-1)$  » renvoient les coordonnées du dernier point de la polyligne.

Il fallait également obtenir une couche avec les coordonnées des bouches de métro et de RER de Paris. Ces données étaient disponibles sur le site [data.ratp.fr](http://data.ratp.fr).

Sur cette première image on observe une très forte concentration de points de départ (en rouge) et d'arrivée (en vert) au niveau de l'entrée Ouest du parc de Vincennes. Cette entrée donne directement sur l'Est de Paris, on peut donc en conclure qu'un certain nombre de coureurs vient de Paris ou en tout cas de l'est de Paris.

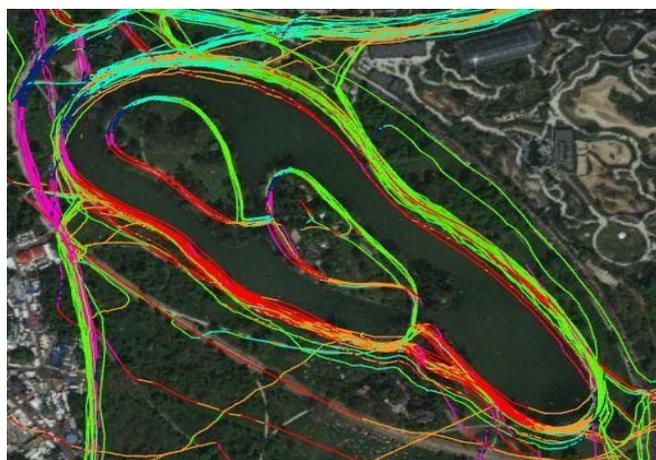
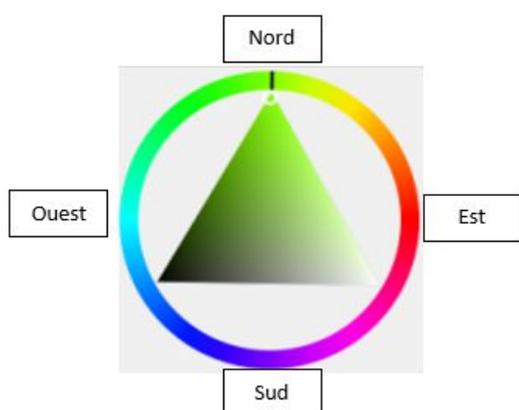


Contre toute attente les points de départ et d'arrivée ne sont pas majoritairement situés au niveau des bouches de métro. Cela peut être dû à deux choses. D'une part les utilisateurs ne commencent à courir qu'au niveau du bois, ce qui pourrait signifier que les personnes ont pour but de courir au bois de Vincennes.

D'autre part, un certain nombre de coureurs partent de chez eux, ce qui pourrait signifier que ce sont les gens proches de Vincennes qui courent dans le parc.

### *Existe-t-il un sens privilégié par les coureurs ?*

Pour déterminer le sens de circulation à Vincennes et à Torcy il fallait connaître l'orientation de chaque morceau de polyligne.



Le

travail a d'abord consisté à découper les traces GPS qui sont des polygones en plusieurs tronçons constitués uniquement de lignes simples. Puis il a fallu recalculer les coordonnées des points de départ et des points d'arrivée de chaque tronçon. Ensuite grâce à ces coordonnées il ne restait plus qu'à calculer l'azimut (qui nous donne un angle en degré entre le Nord et la droite formée par le point de départ et d'arrivée) de chaque ligne simple par le biais d'une formule dans la calculatrice de champ de QGIS «  $degree(azimuth(make\_point(xat(0),yat(0)),make\_point(xat(-1),yat(-1))))$  ».

Une fois l'orientation de chaque ligne déterminée, il s'agissait de trouver un moyen de représentation du sens de circulation. La meilleure option était de mettre une couleur différente selon l'orientation de chaque tronçon.

Sur cette carte on observe assez bien la tendance de circulation autour du lac Daumesnil à l'Ouest du parc de Vincennes : les coureurs courent dans le sens inverse des aiguilles d'une montre, c'est-à-dire qu'au nord du lac les coureurs se dirigent vers Paris et au sud vers Marne la Vallée.

## Conclusion et perspectives

---

### Retour des travaux réalisés

La demande du commanditaire était la réalisation de cartes sur les zones du bois de Vincennes. Ces cartes devaient aider à la compréhension du comportement des coureurs et à l'analyse de l'utilisation du territoire par ces derniers. Ces analyses devaient permettre de faciliter l'aménagement du territoire pour favoriser l'accessibilité des principales zones de courses.

Pour répondre à cette demande plusieurs cartes ont été effectuées :

- une carte sur les points de départ et d'arrivée
- une carte sur le sens de circulation

Toutes les cartes permettent d'émettre des hypothèses sur le comportement des coureurs et sur l'utilisation du territoire. Elles répondent donc bien au sujet posé par le commanditaire. Elles pourront à l'avenir servir pour des analyses plus poussées sur le sujet.

### Perspectives

Avec plus de temps on aurait pu approfondir l'analyse des données Strava en créant d'autres cartes notamment une sur l'altitude des traces GPS, une autre sur les traces GPS aux différentes périodes de la journée, matin, soir etc. On aurait pu également créer des cartes en combinant les différentes informations extraites durant ce projet comme par exemple la longueur des courses et la période de la semaine. Toutes les données shapefiles pourront être utilisées pour une analyse plus poussée du comportement des coureurs et de l'utilisation du territoire.