
Data Analysis

Project - Velib



Doan-Kien THAI, Jonas MELANDSOR, Elsa DANH-NGHET
Group 12, 4MA, 2024

INSA
TOULOUSE

Table of contents

01

Introduction to
the data

02

Principal
component
analysis

03

Clustering

04

Multiple
correspondence
analysis

05

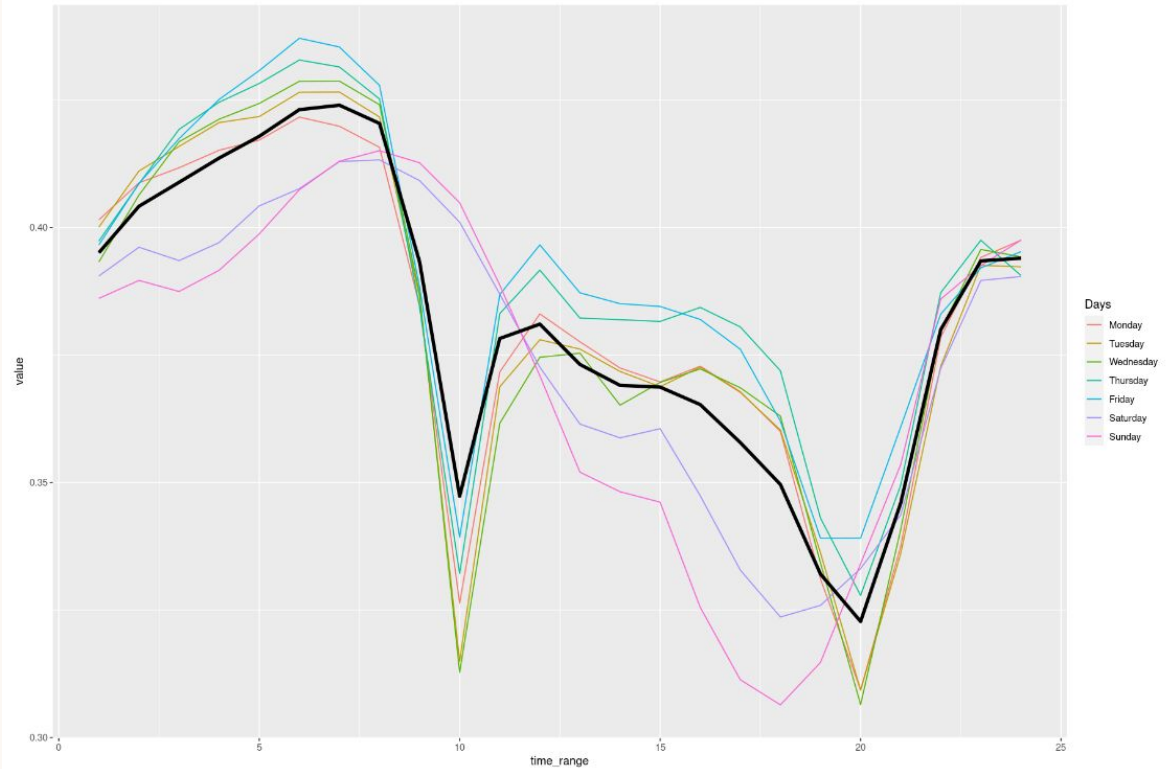
Conclusion

01

Introduction To the data



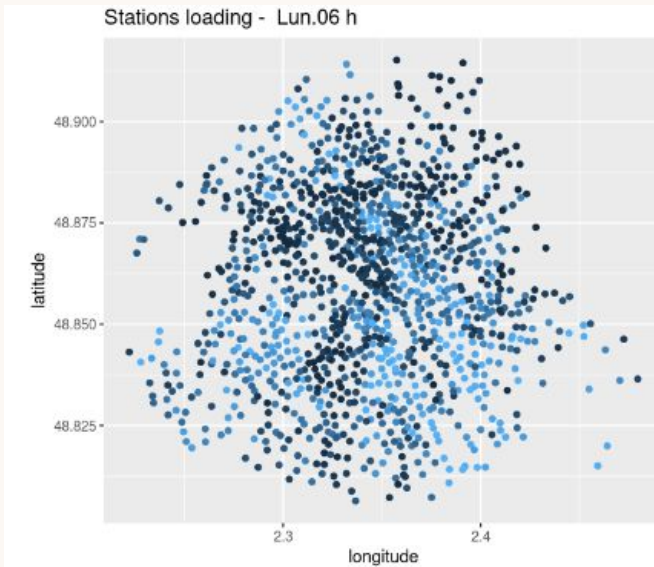
General



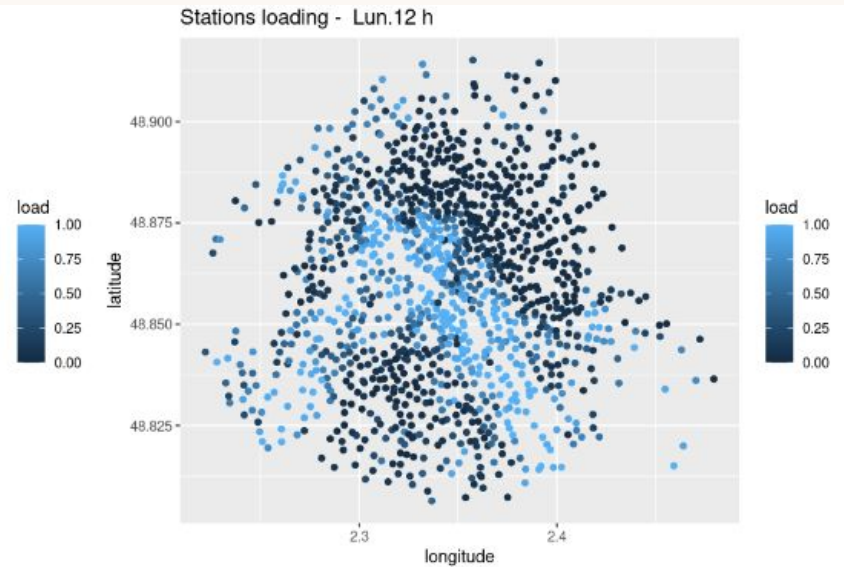
- **Least loaded station on average :**
Hornet (Bagnolet), 0.0161
- **Fullest loaded station on average :**
Insurrection aout 1944 (Ivry), 0.9194
- **Repeating pattern**

Loading of the stations in Paris

Monday, 6am



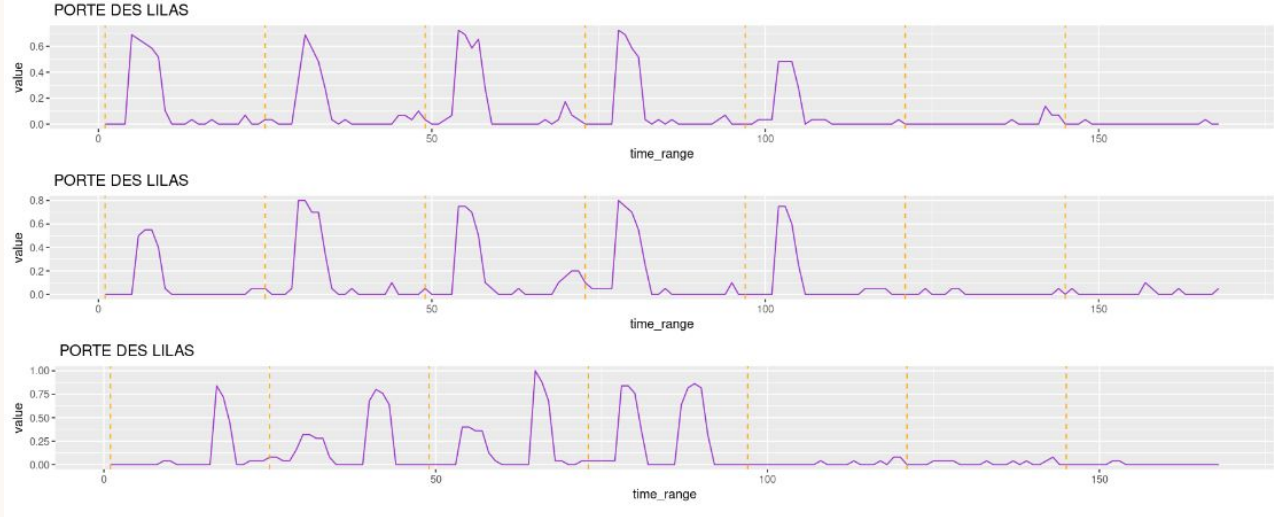
Monday, 12am





Multiple occurrences

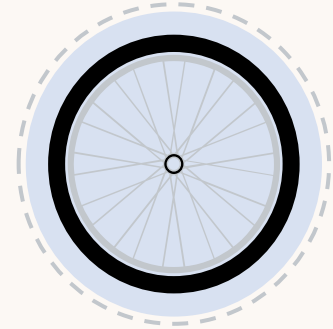
Porte des lilas



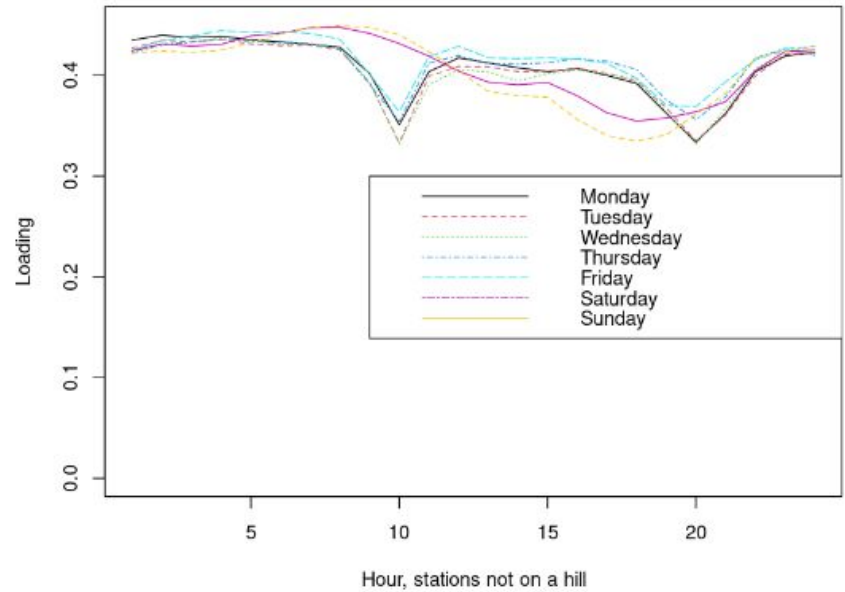
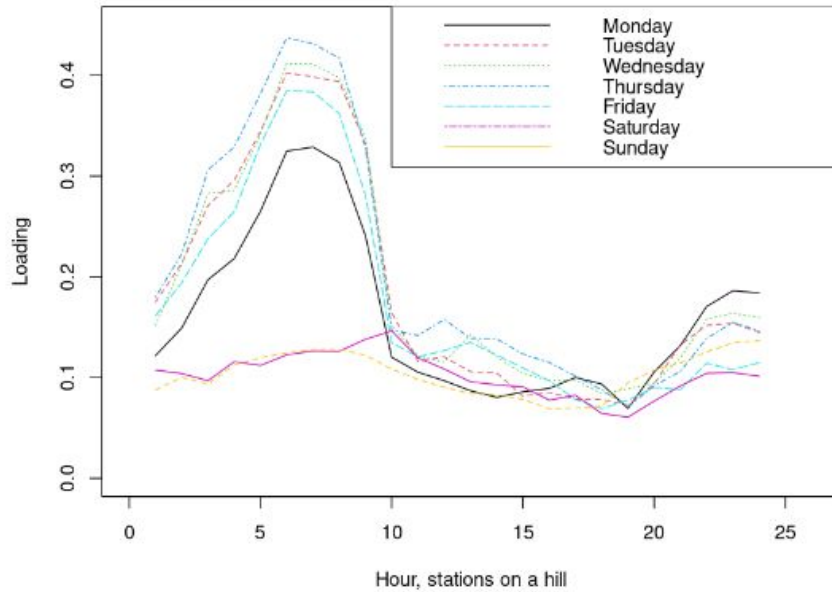
Clichy

Average loading of Clichy: 0.0997

Average loading : 0.3816



Impact of being on a hill

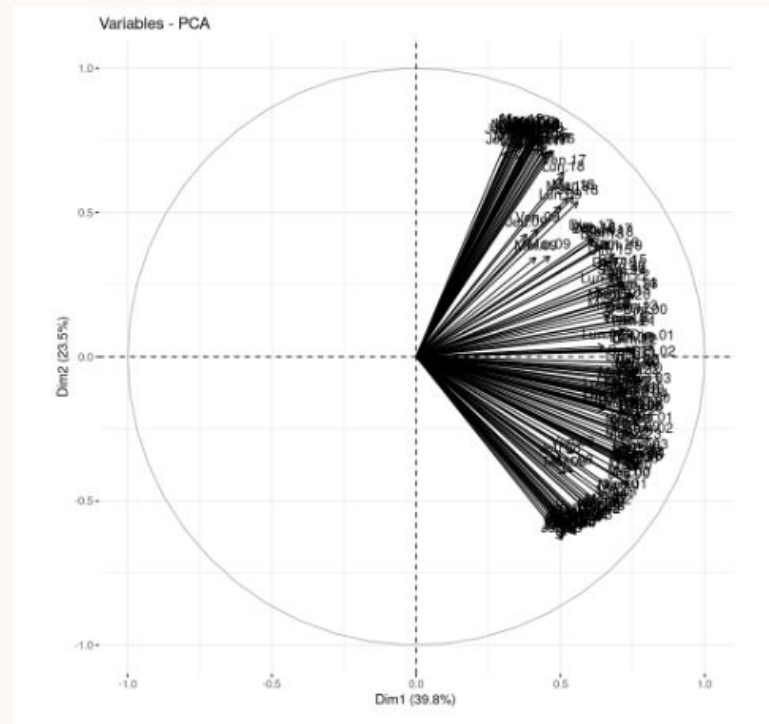


Principal Component Analysis

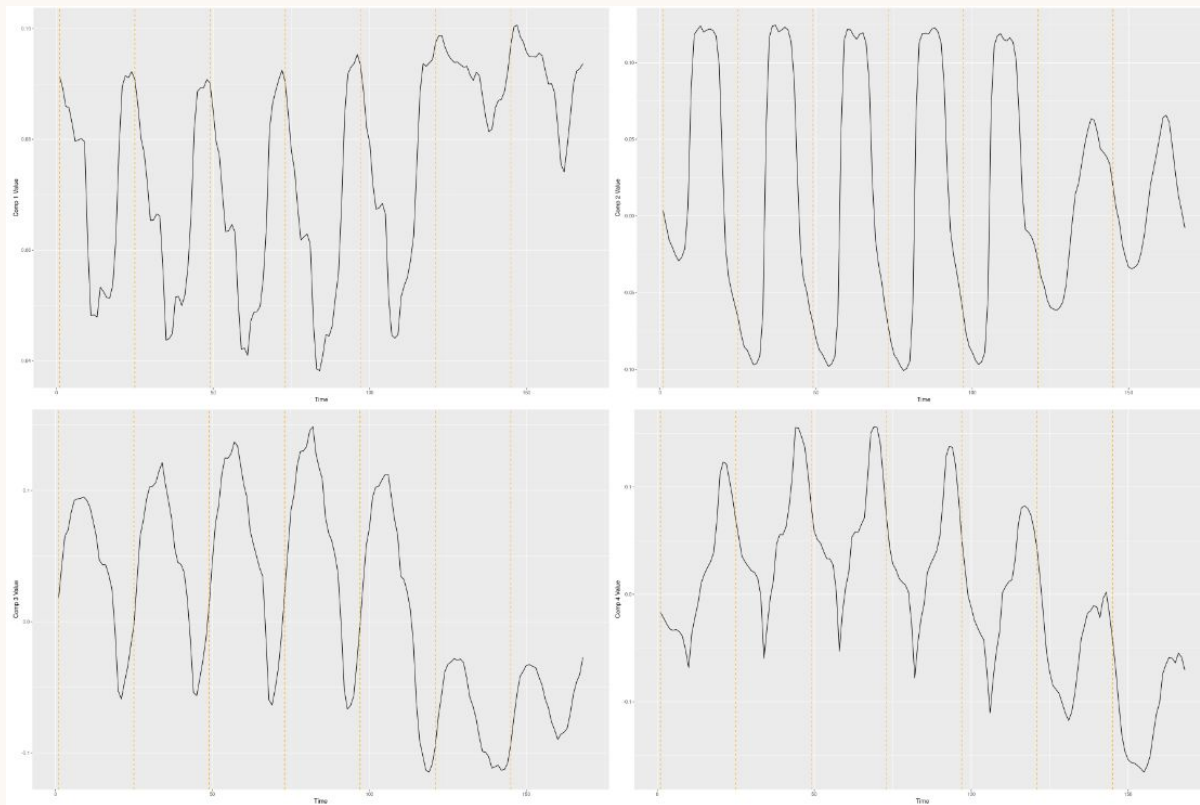


Choice of the number of components

eigenvalue		cumulative percentage of variance	
6 × 3 of			
comp 1	66.888963		39.81486
comp 2	39.486446		63.31870
comp 3	8.867649		68.59706
comp 4	7.305326		72.94547
comp 5	5.503113		76.22113
comp 6	3.245742		78.15312



Interpretation of the components



1st: Stations prominently losing bikes during the working day

2nd: Stations prominently receiving bikes during the working day

3rd: Hill stations

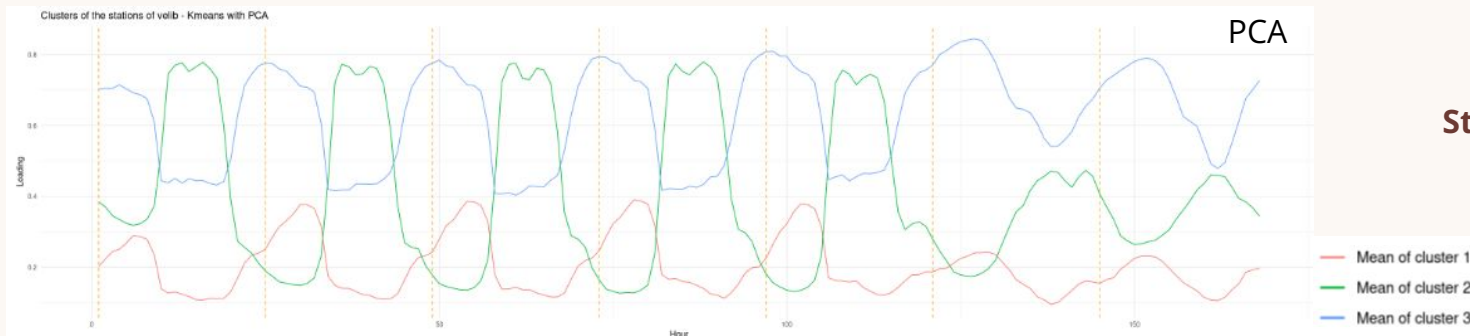
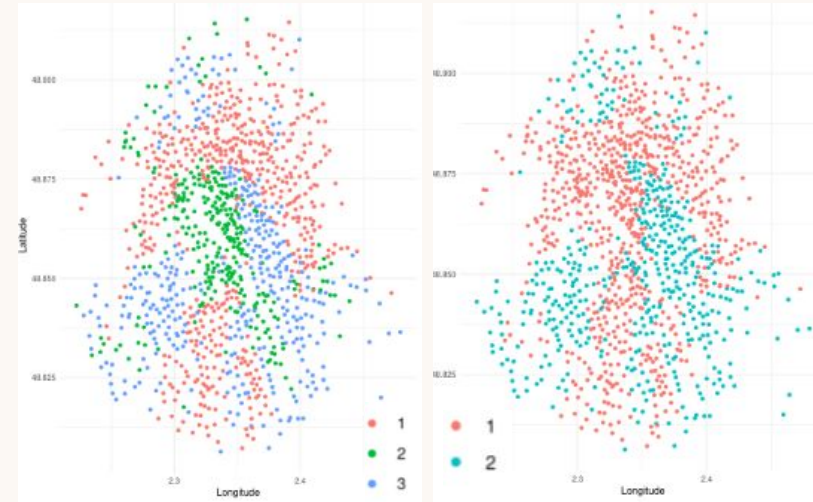
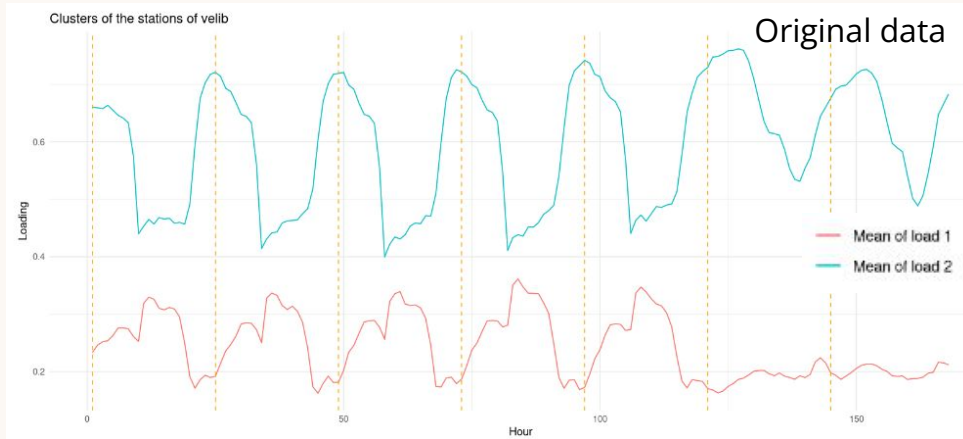
4th: Redistribution of bikes

03

Clustering

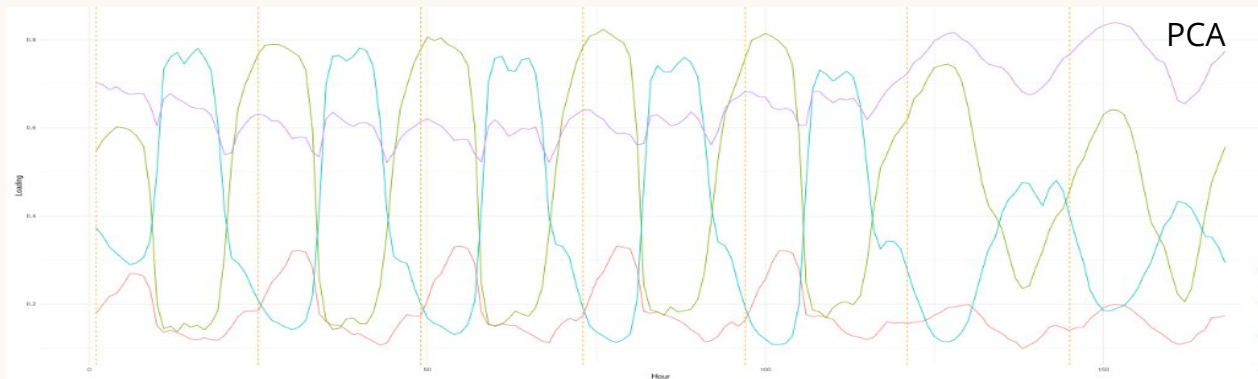
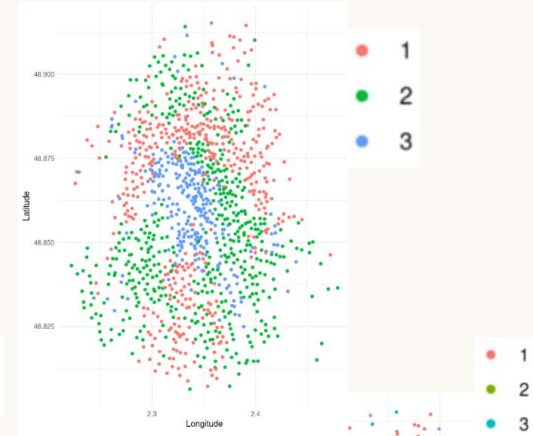
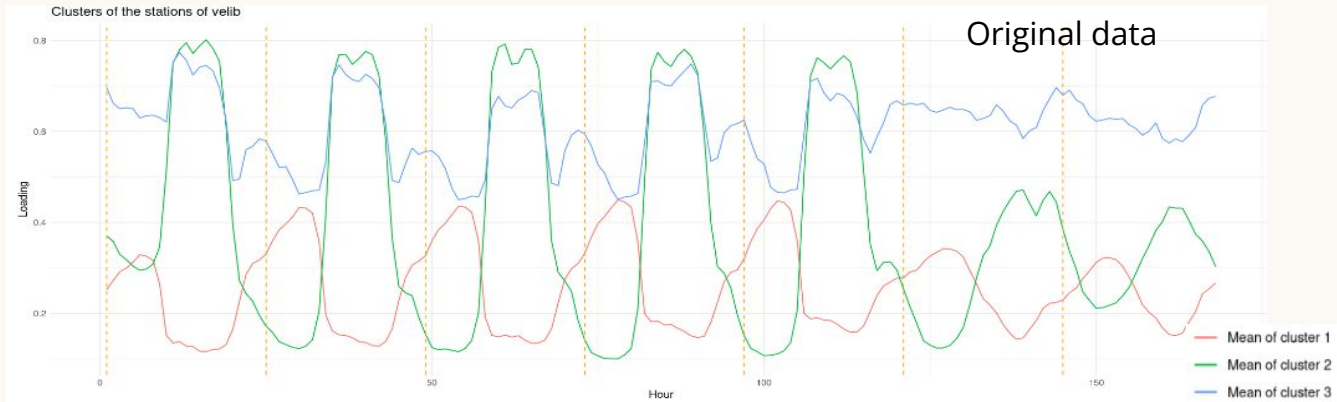


Kmeans



**Stations on a hill :
red line**

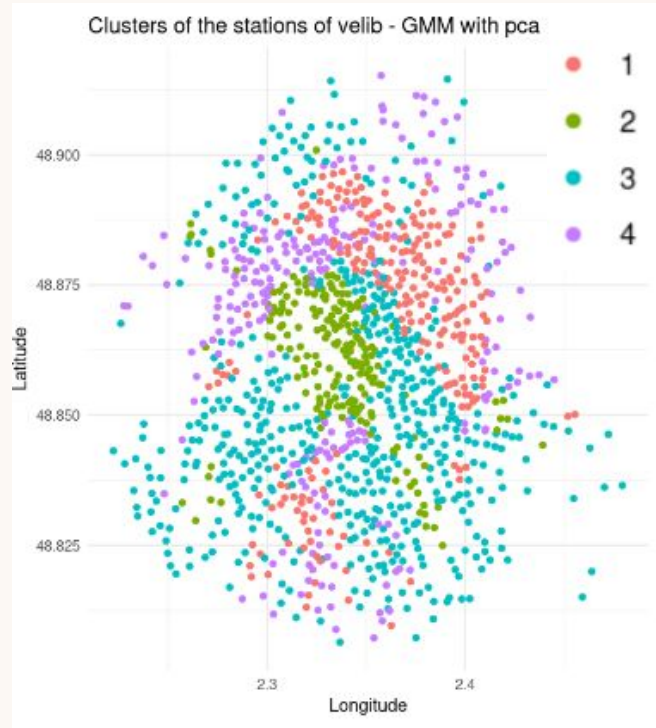
Hierarchical Agglomerative Clustering



Stations on a hill : red line



Gaussian Mixture Models

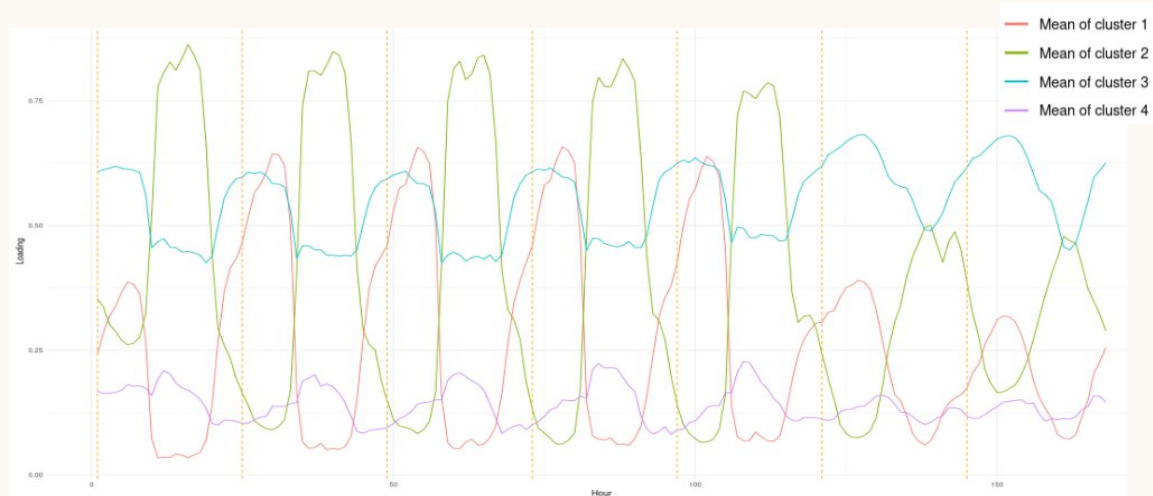


VVV cov. matrix

Not hill ->

Hill ->

	1	2	3	4
0	184	192	496	190
1	63	0	13	51



Conclusion for quantitative clustering

Clusters:

- Kmeans: 2
- AHC: 3
- Kmeans (PCA): 3
- AHC (PCA): 4
- GMM (PCA): 4

Assigning of hill stations:

- Kmeans: Same
- AHC: Same
- Kmeans (PCA): Same
- AHC (PCA): Same
- GMM (PCA): Three different

ARI	Kmeans	AHC	Kmeans (PCA)	AHC (PCA)	GMM (PCA)
Kmeans	1	0.7947	0.5458	-	-
AHC		1	-	1	-
Kmeans (PCA)			1	0.7192	0.5058
AHC (PCA)				1	0.2953
GMM (PCA)					1

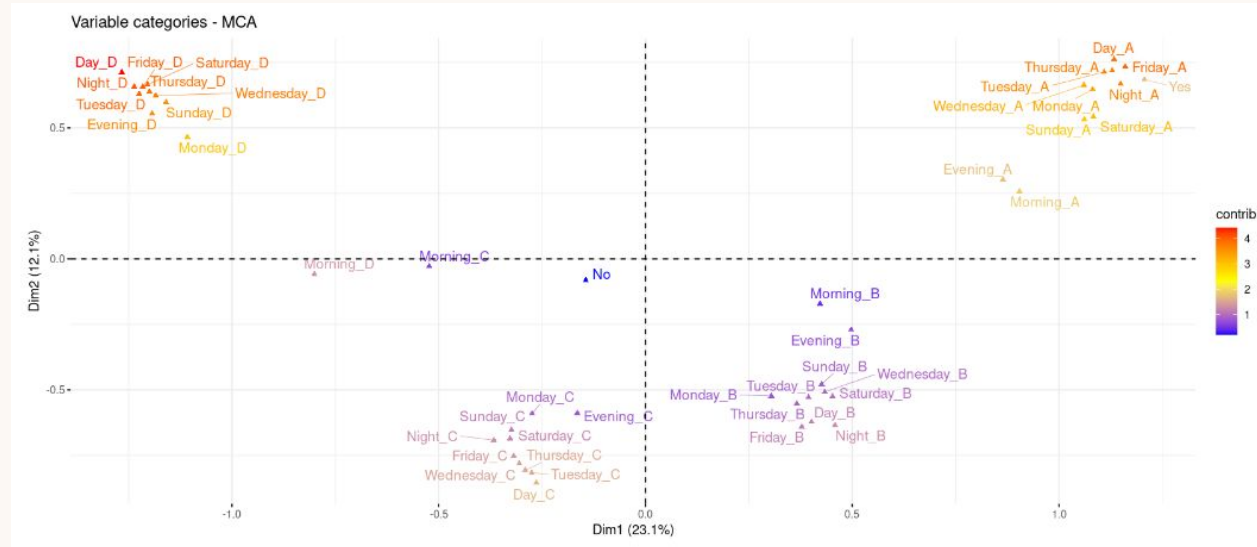
Multiple Correspondance Analysis



First dataframe

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Night	Morning	Day	Evening	Hill
1	A	B	C	B	C	B	B	C	A	B	C	No
2	B	D	C	D	C	D	C	D	A	C	D	No

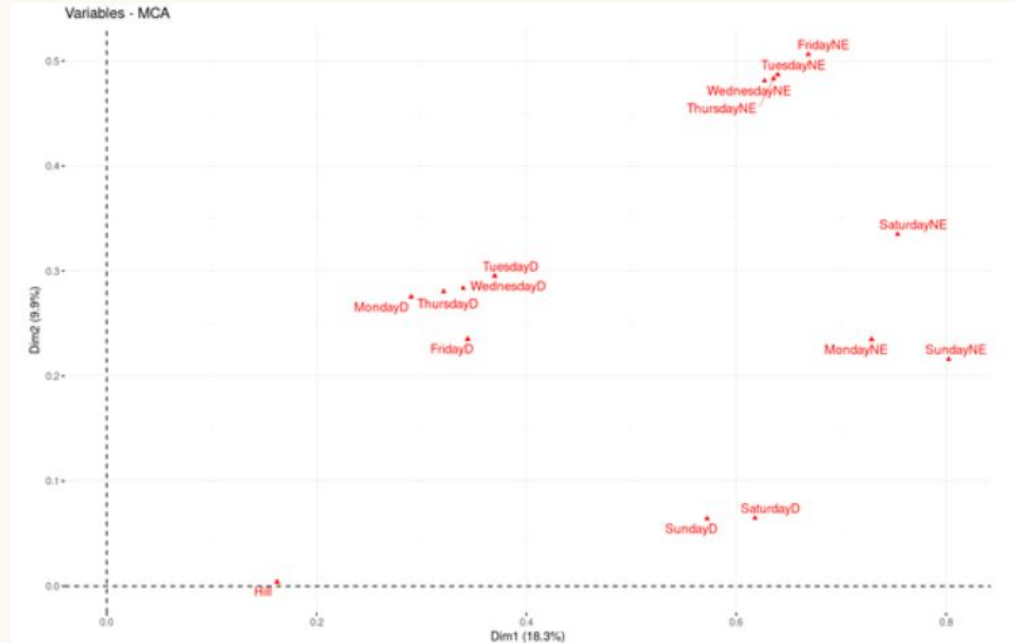
- A-B-C-D : Categories created from the quartiles of each variable
- First dimension : loading of the stations
- Being on a hill : impact on the loading



Second dataframe

	MondayNE	MondayD	TuesdayNE	TuesdayD	WednesdayNE	WednesdayD	ThursdayNE	ThursdayD	FridayNE	FridayD	SaturdayNE	SaturdayD	SundayNE	SundayD	Hill
1	B	A	C	B	C	A	C	B	C	B	C	B	B	B	No
2	D	B	D	B	D	B	D	B	D	B	D	C	D	C	No

- First dimension : Loading
- Second dimension: Day/Night
- Weekdays and weekend still separated



05

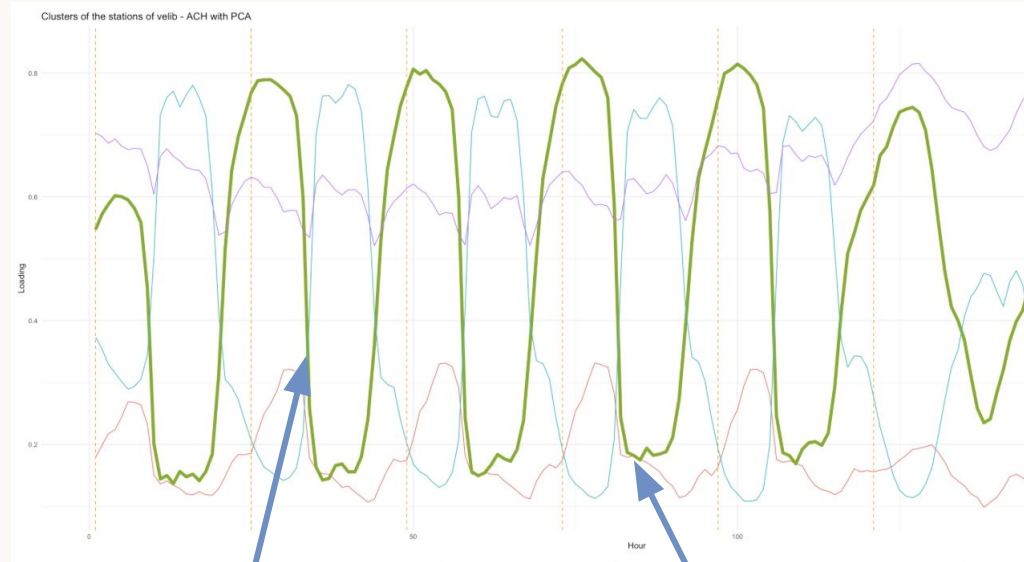
Conclusion



Where to open new stations?



<https://www.southcoasttoday.com/story/lifestyle/2007/07/15/quickie-bike-rentals-make-it/52847851007/>



Steep downfall
around 8-10am

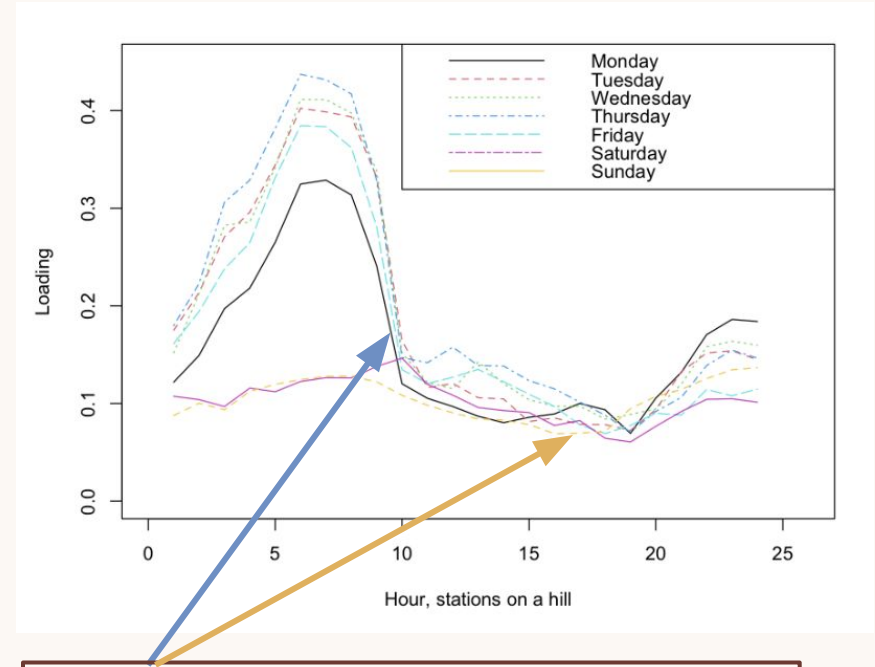
Low loading during the day

How can bikes be redistributed?



<https://www.leparisien.fr/info-paris-ile-de-france-oise/transports/la-ville-de-paris-recupere-les-velib-egares-24-04-2018-7681821.php>

Bikes can be moved using special trucks



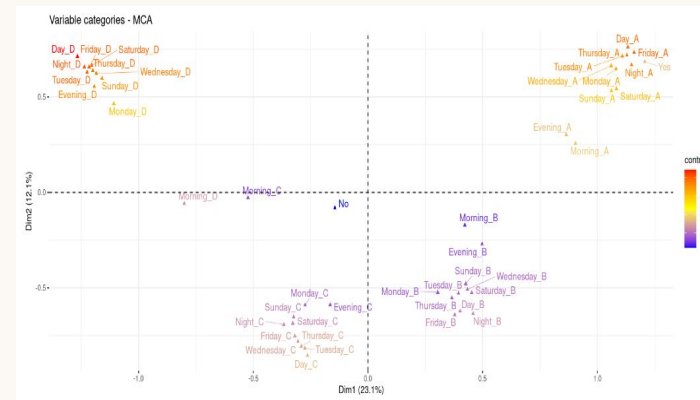
Steep downfall during 8-10am but not much recovery in the afternoon

Self-distribute by users

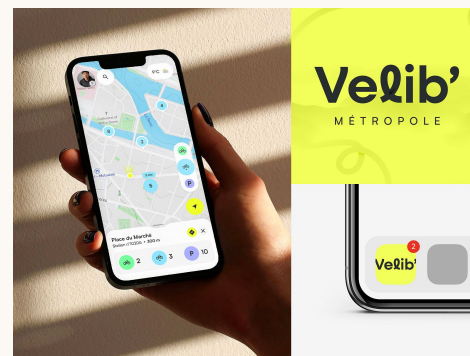


<https://www.voltage.fr/news/velib-rendre-son-velo-dans-une-station-pleine-n-est-plus-un-probleme-61201>

Users redistribute bikes to high-altitude stations to get extra 15m-ride (Vélib, 2014)



By using MCA, we can further mark stations that are often empty



Encouraging redistributing to empty stations



Thanks!

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)