



Sciences Economiques & Sociales de la Santé
& Traitement de l'Information Médicale

sesstim.univ-amu.fr

Christophe GENOLINI

*MCU - Université Paris Nanterre
Fondateur Start-up R++ The Next Step*

**CopyMean :
une nouvelle méthode d'imputation des données longitudinales**

novembre 2017



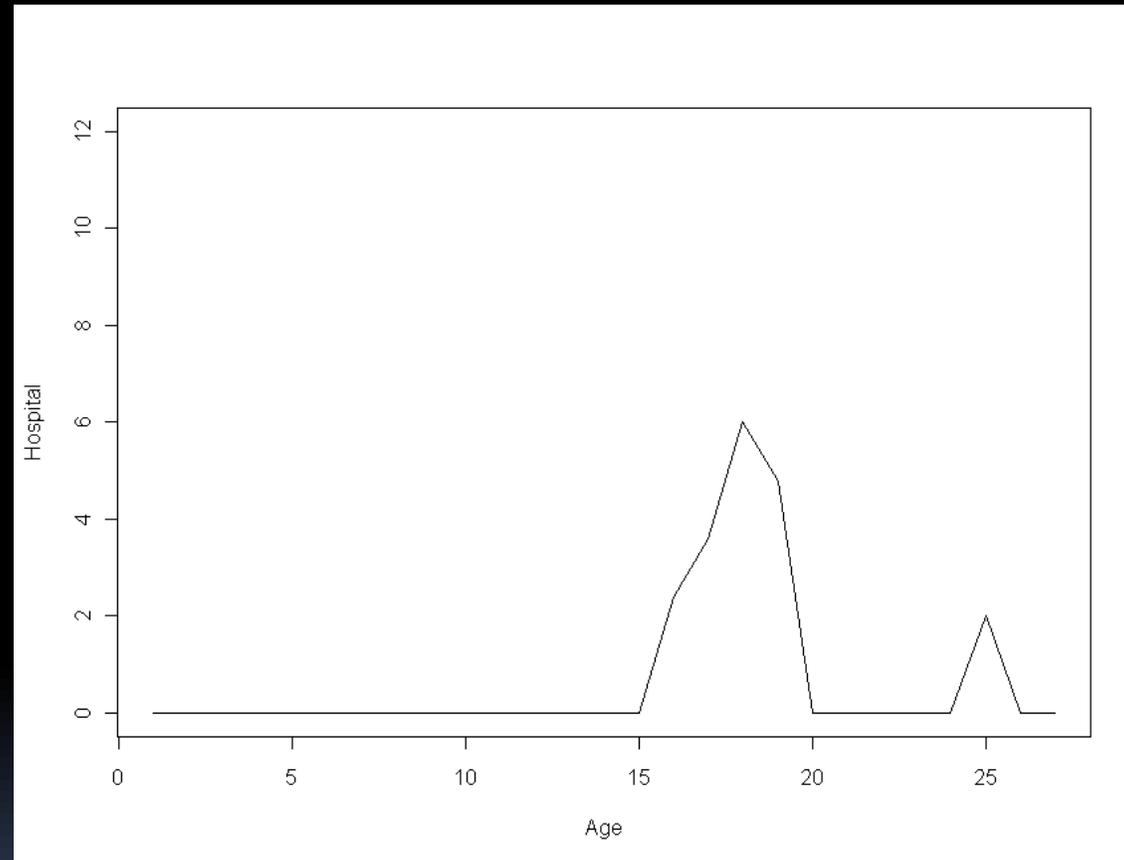
[Cliquez ici pour voir l'intégralité des ressources associées à ce document](#)



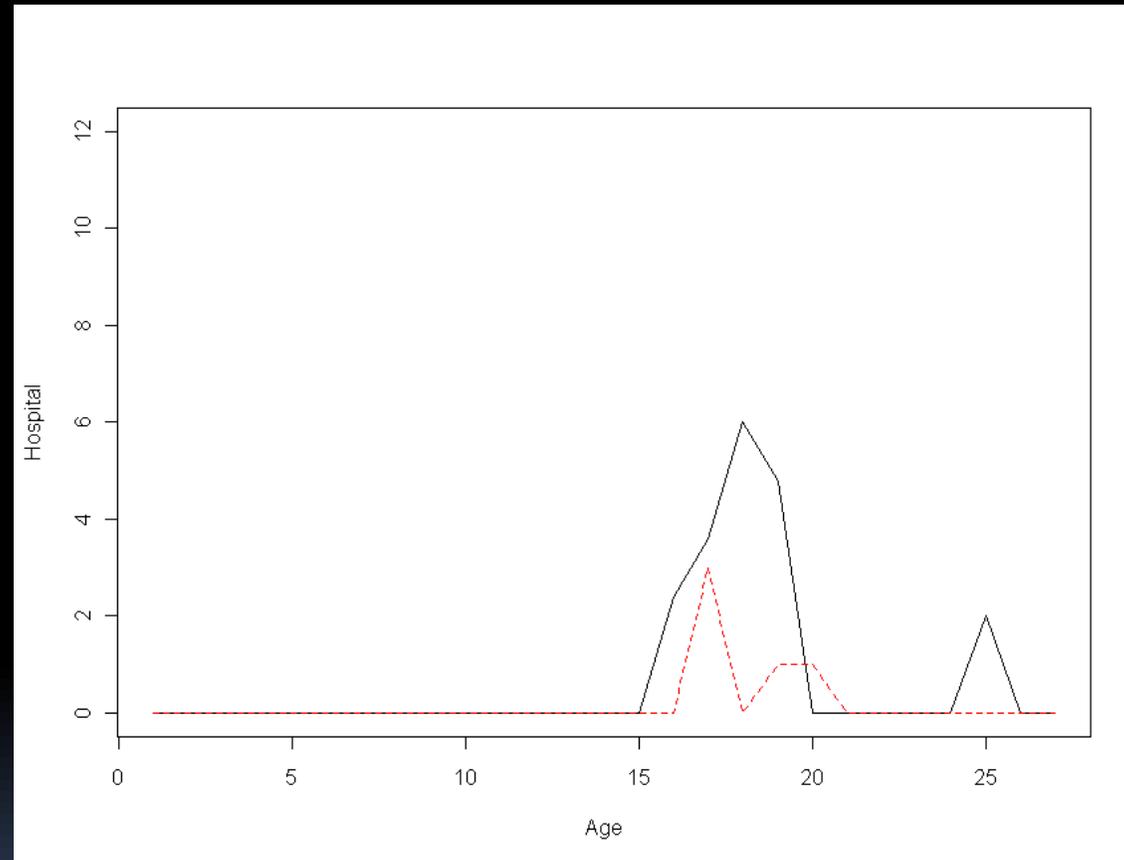
Christophe Genolini & René Écochard

COPYMEAN :
IMPUTATION DES DONNÉES LONGITUDINALES

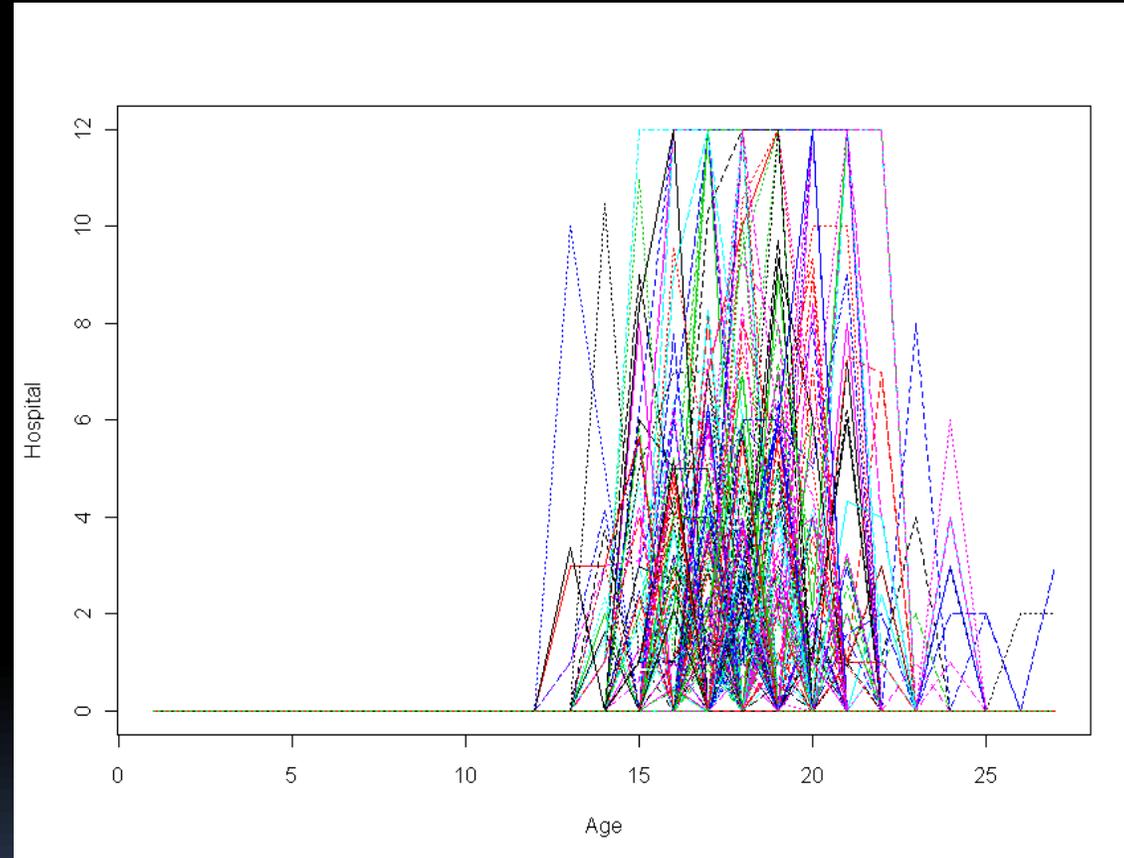
Une trajectoire

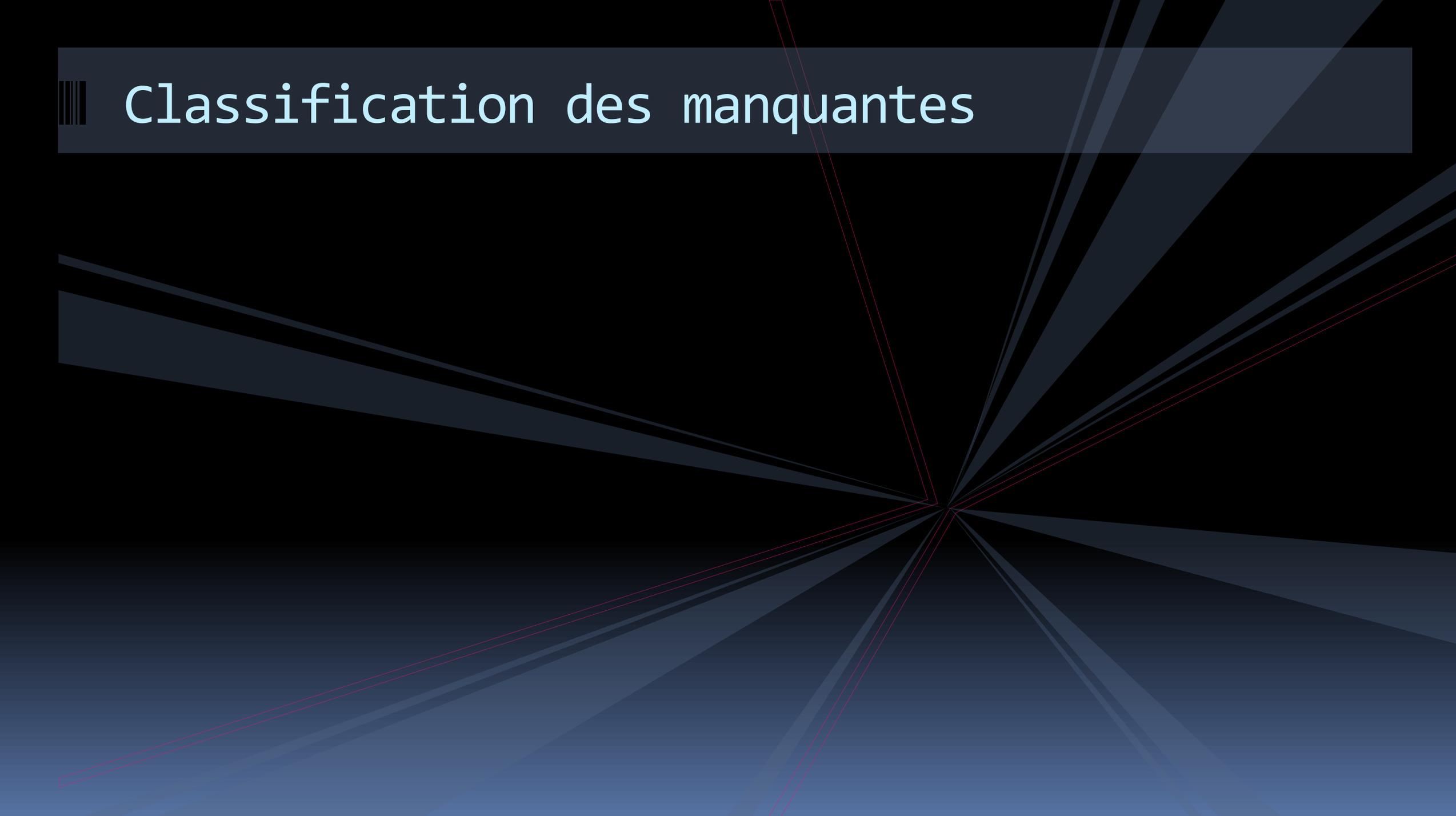


Deux trajectoires



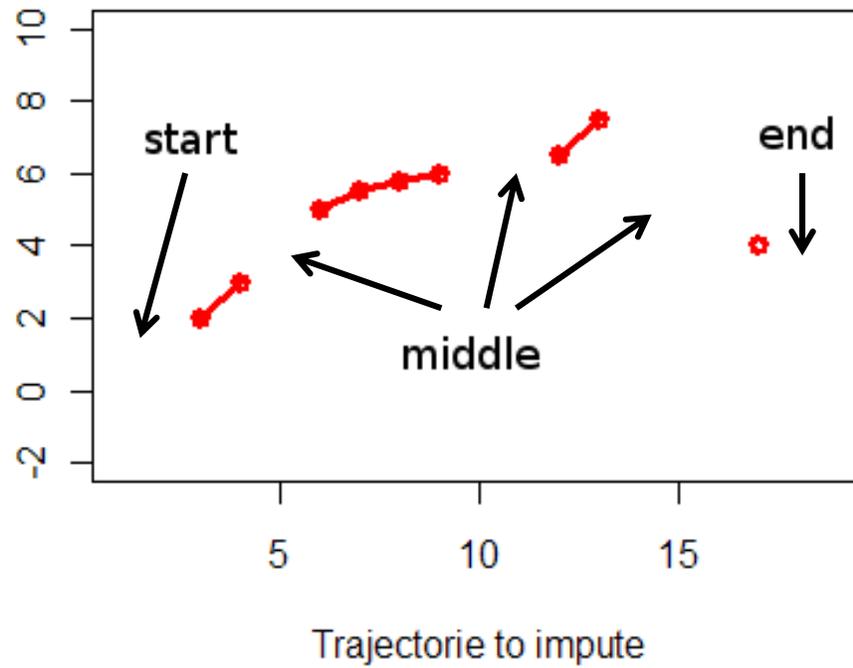
Données longitudinales



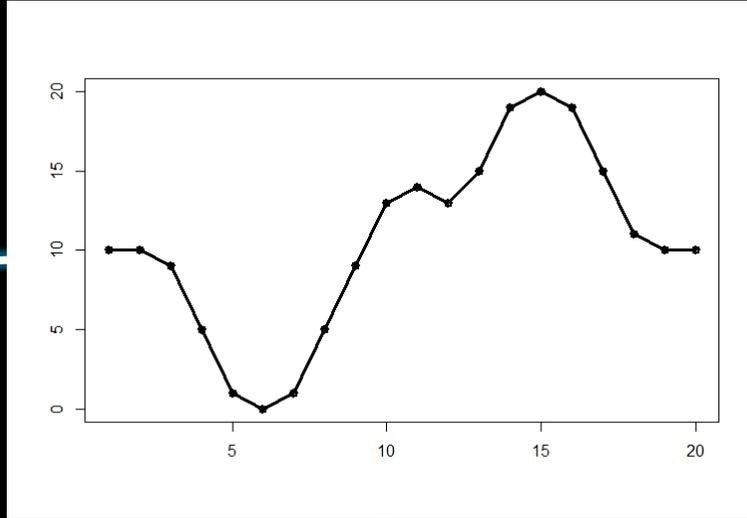


Classification des manquantes

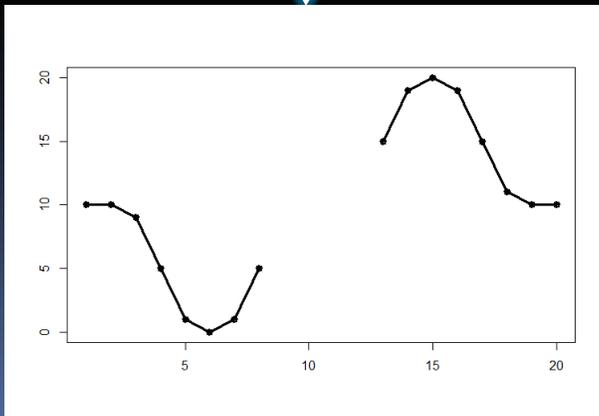
Position



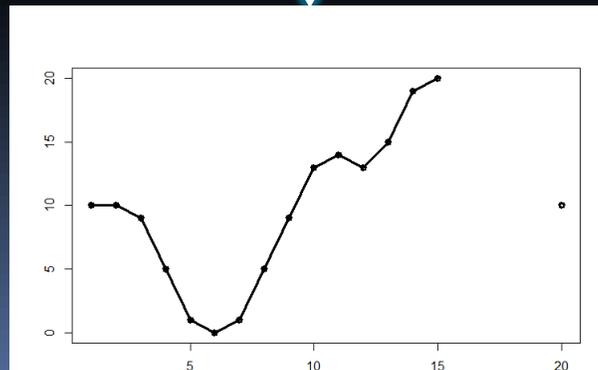
Cause



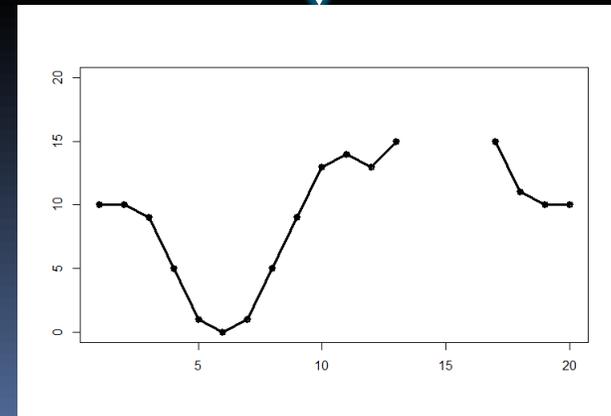
MCAR



MAR

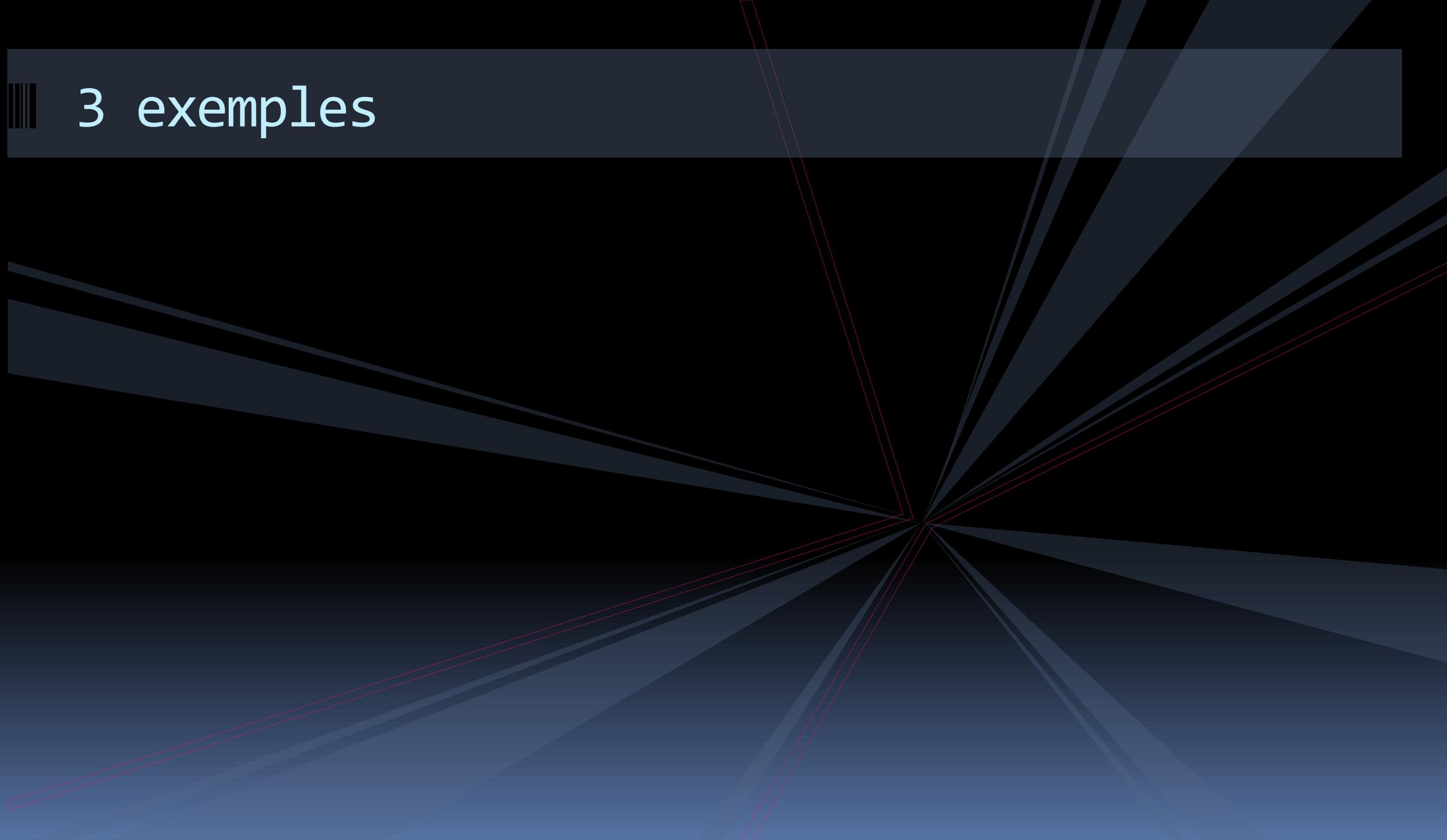


MNAR

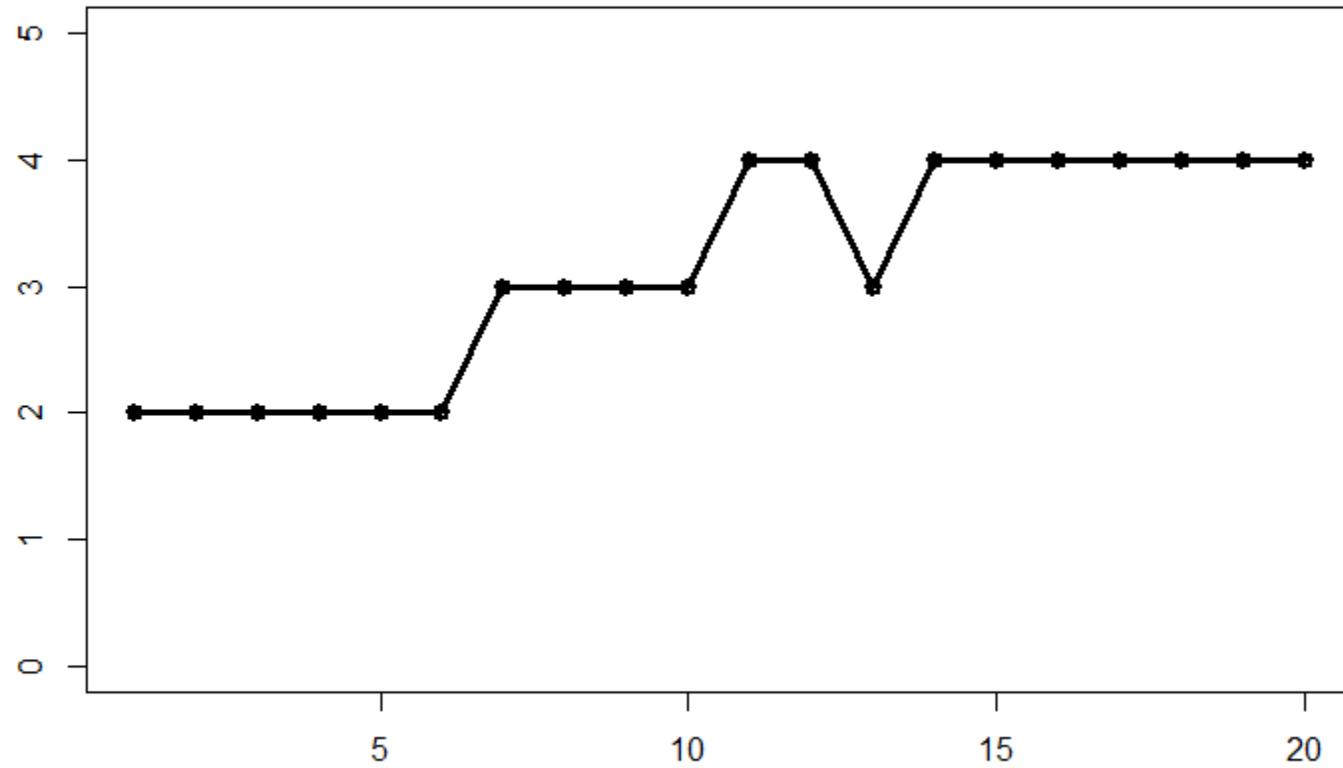




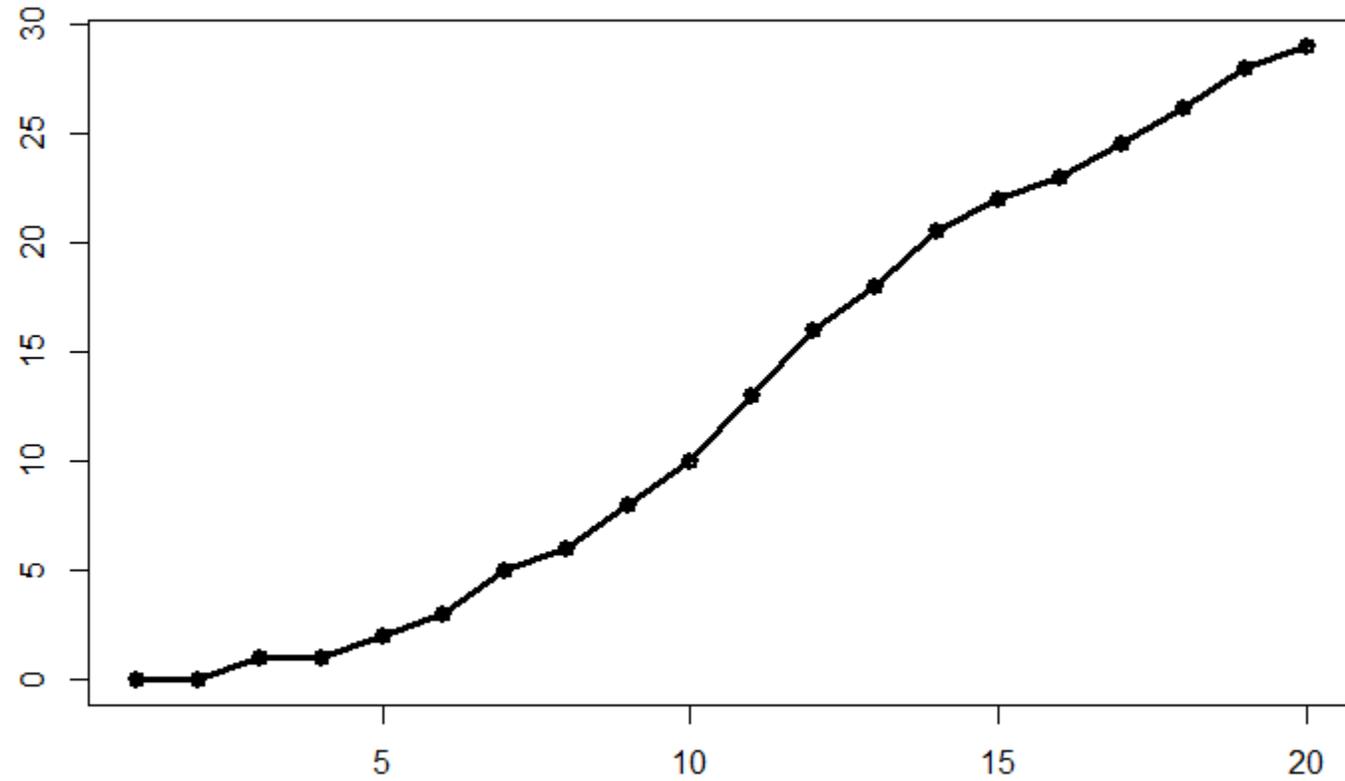
3 examples



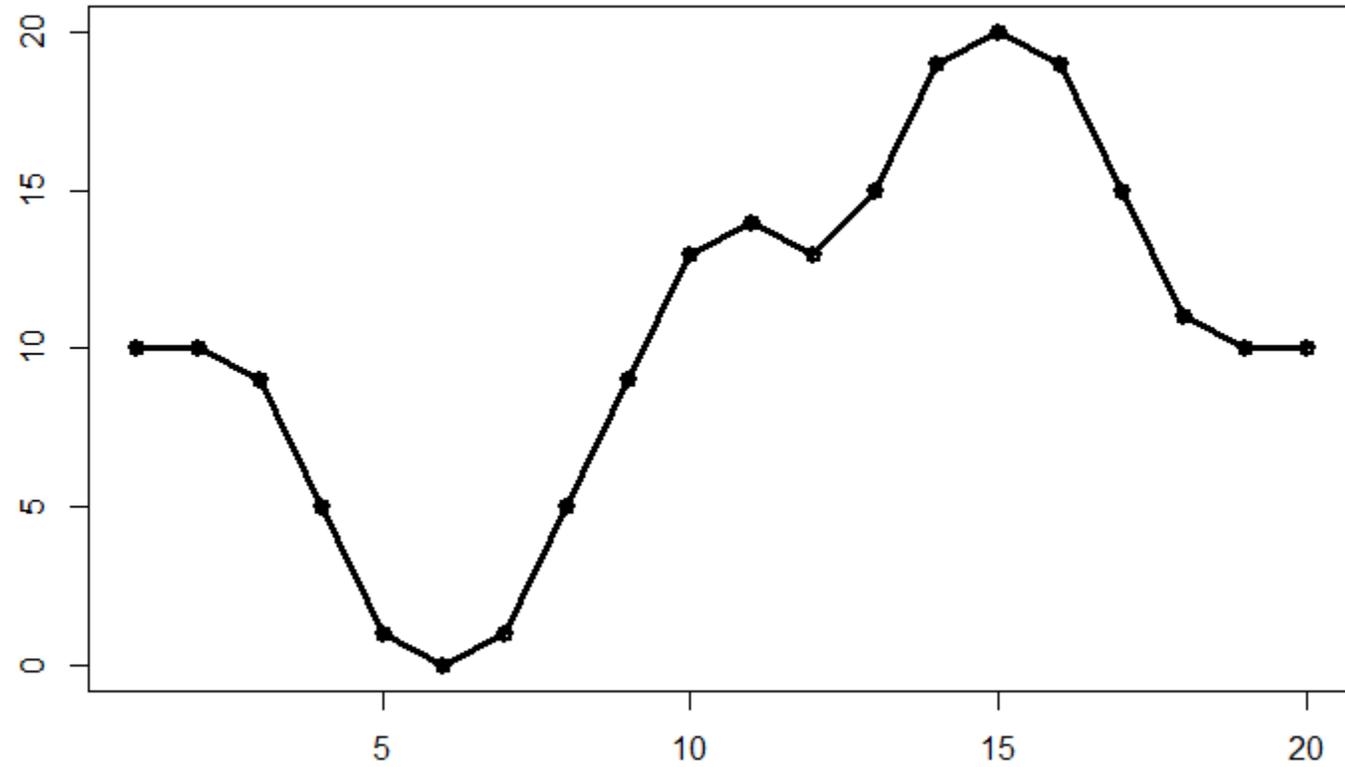
Exemple 1



Exemple 2

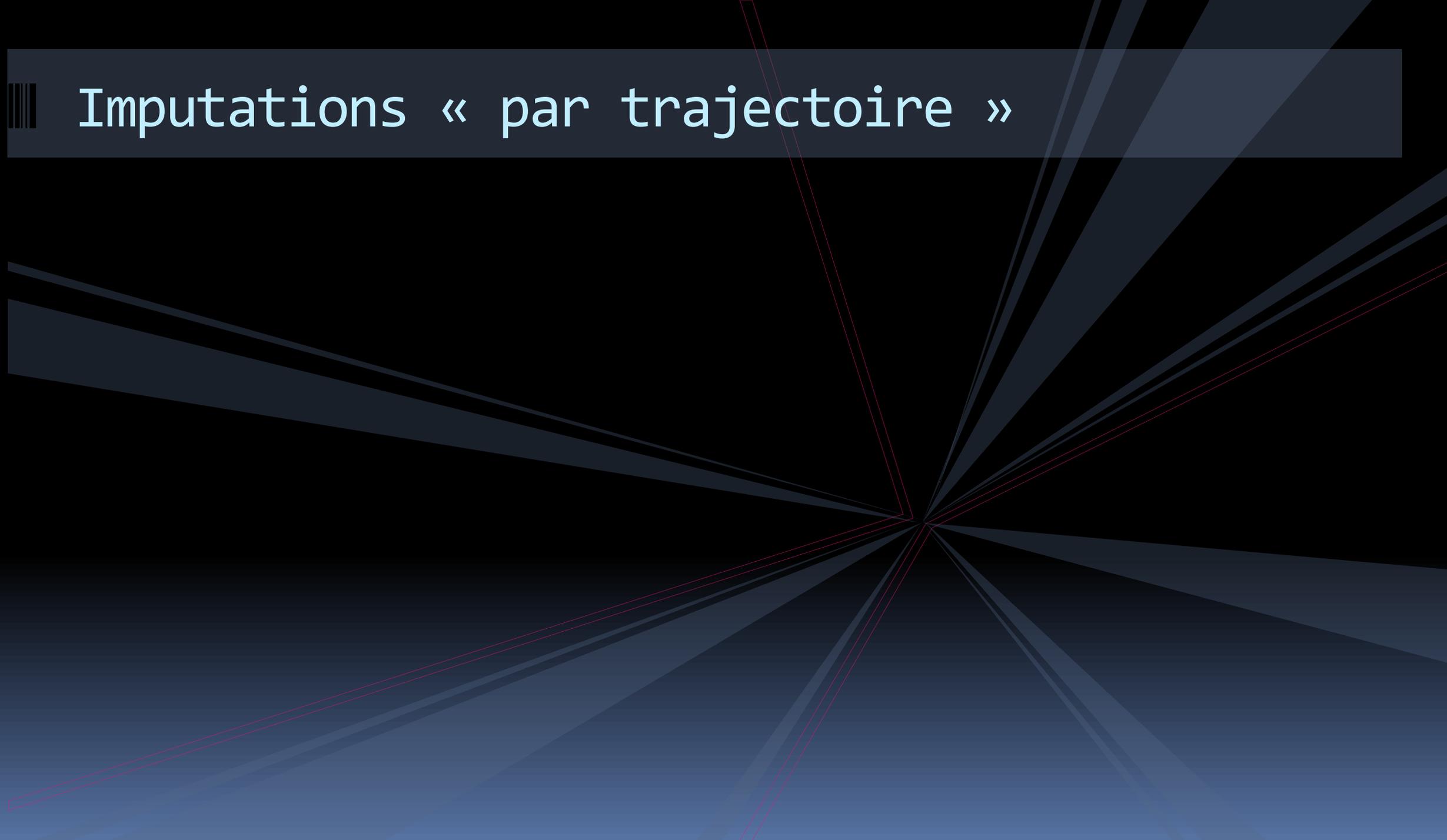


Exemple 3

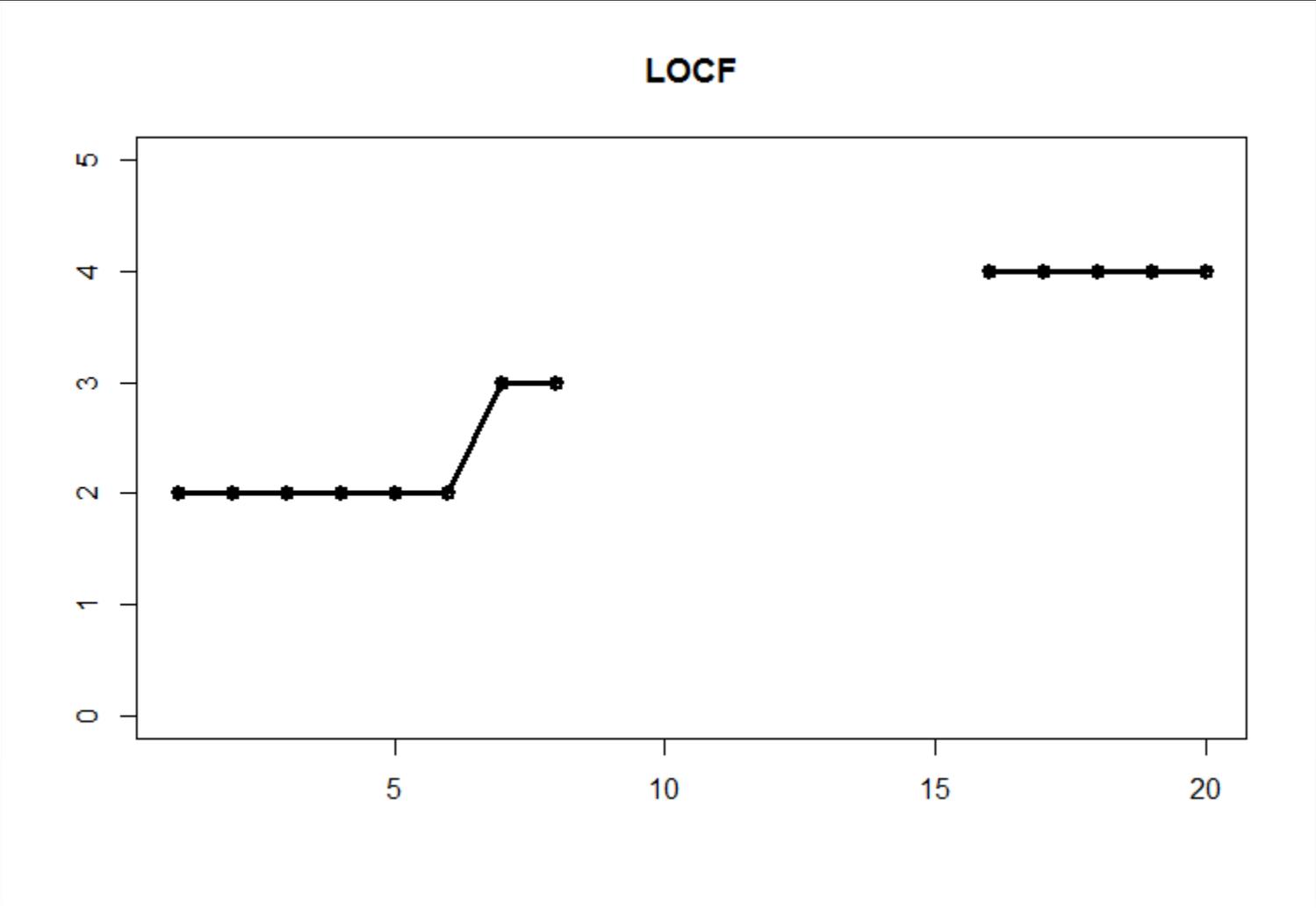




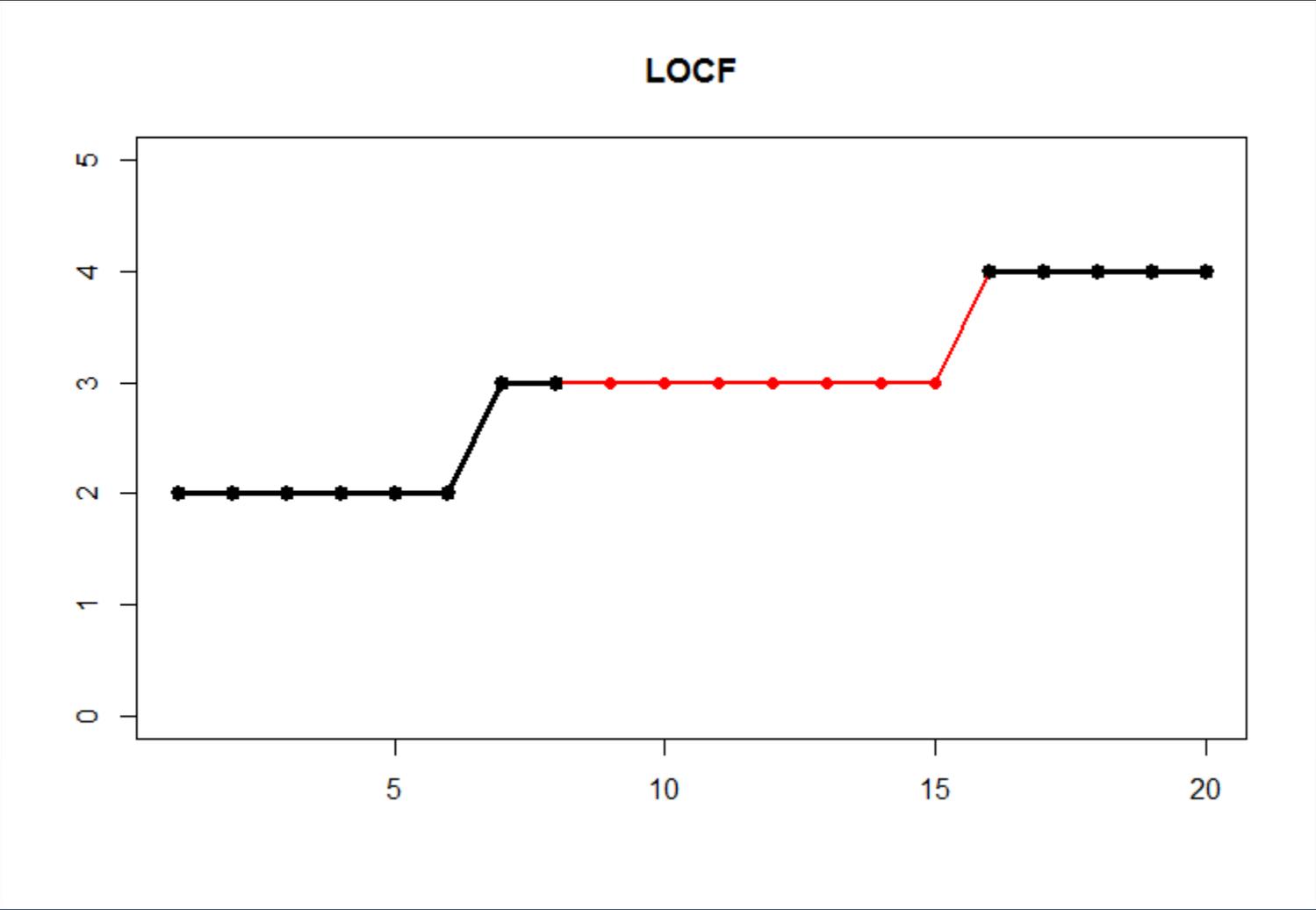
Imputations « par trajectoire »



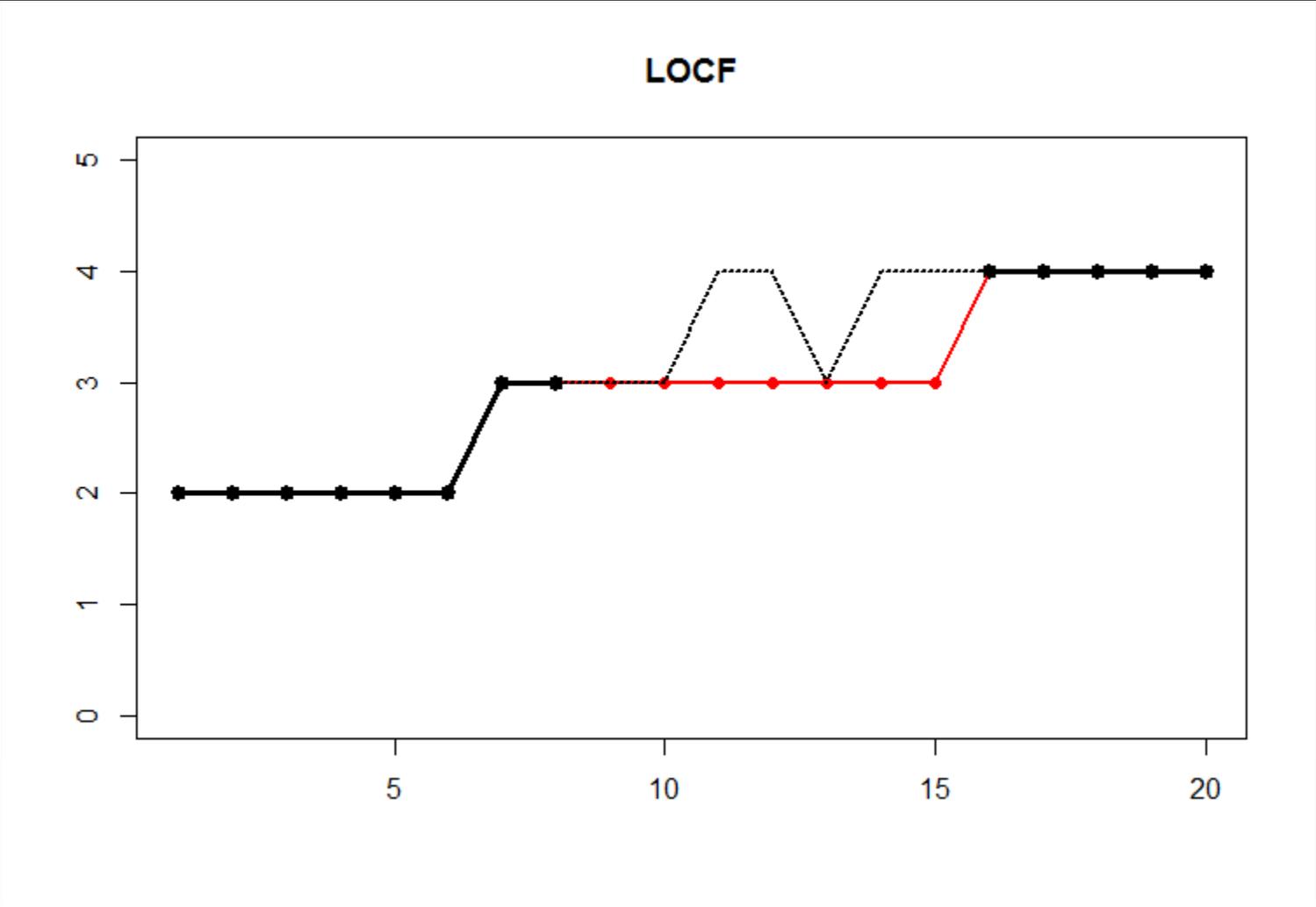
LOCF



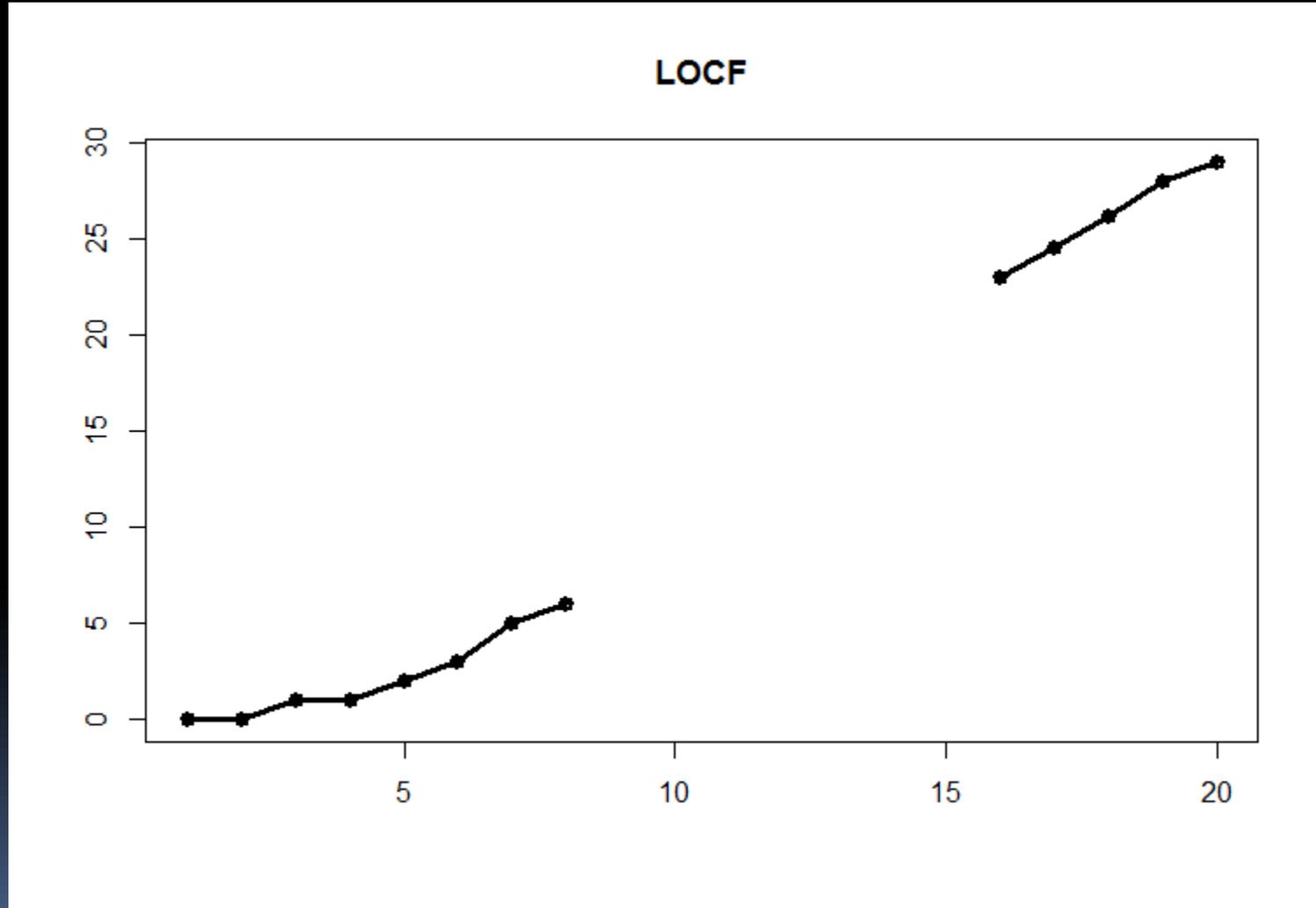
LOCF



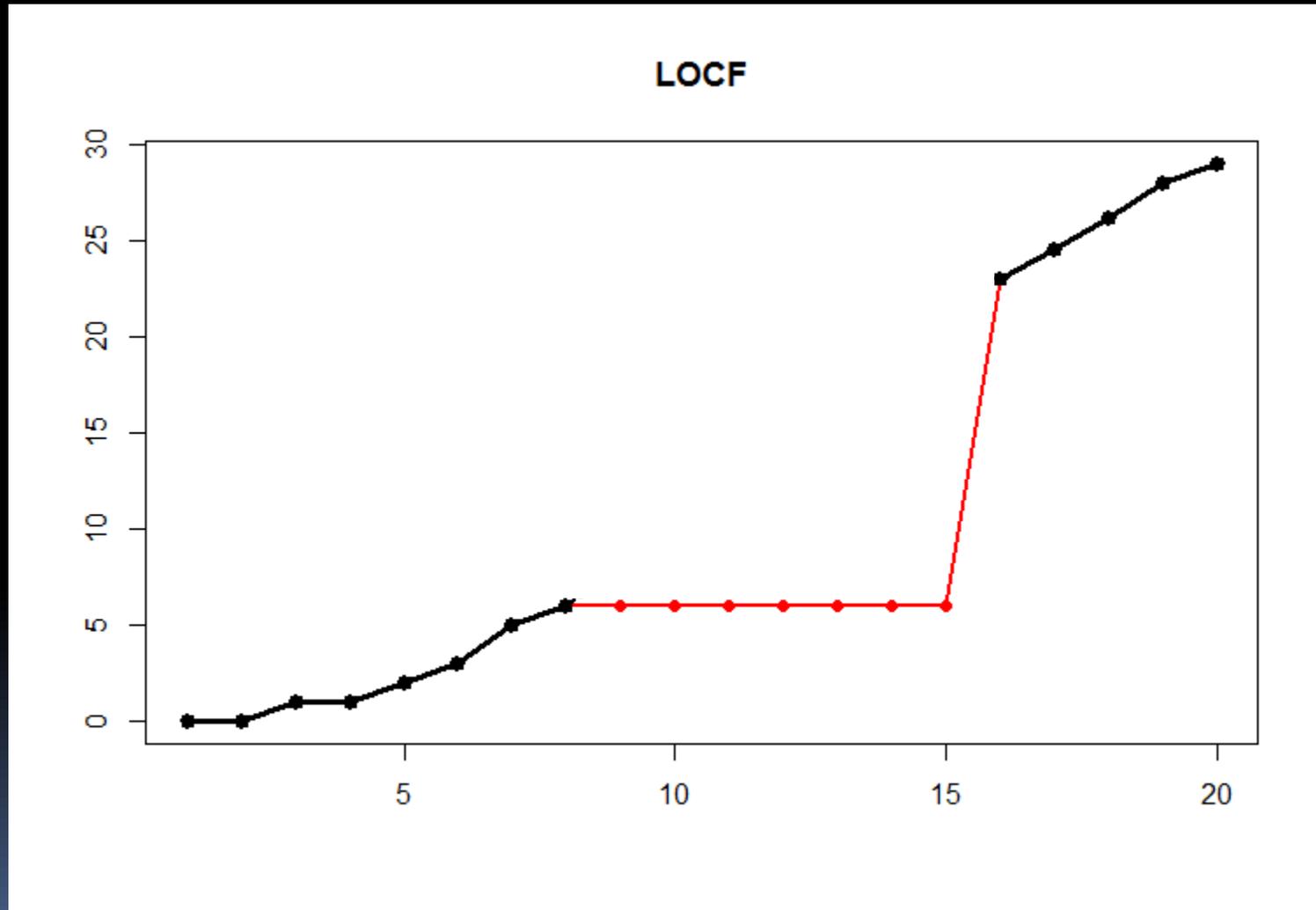
LOCF



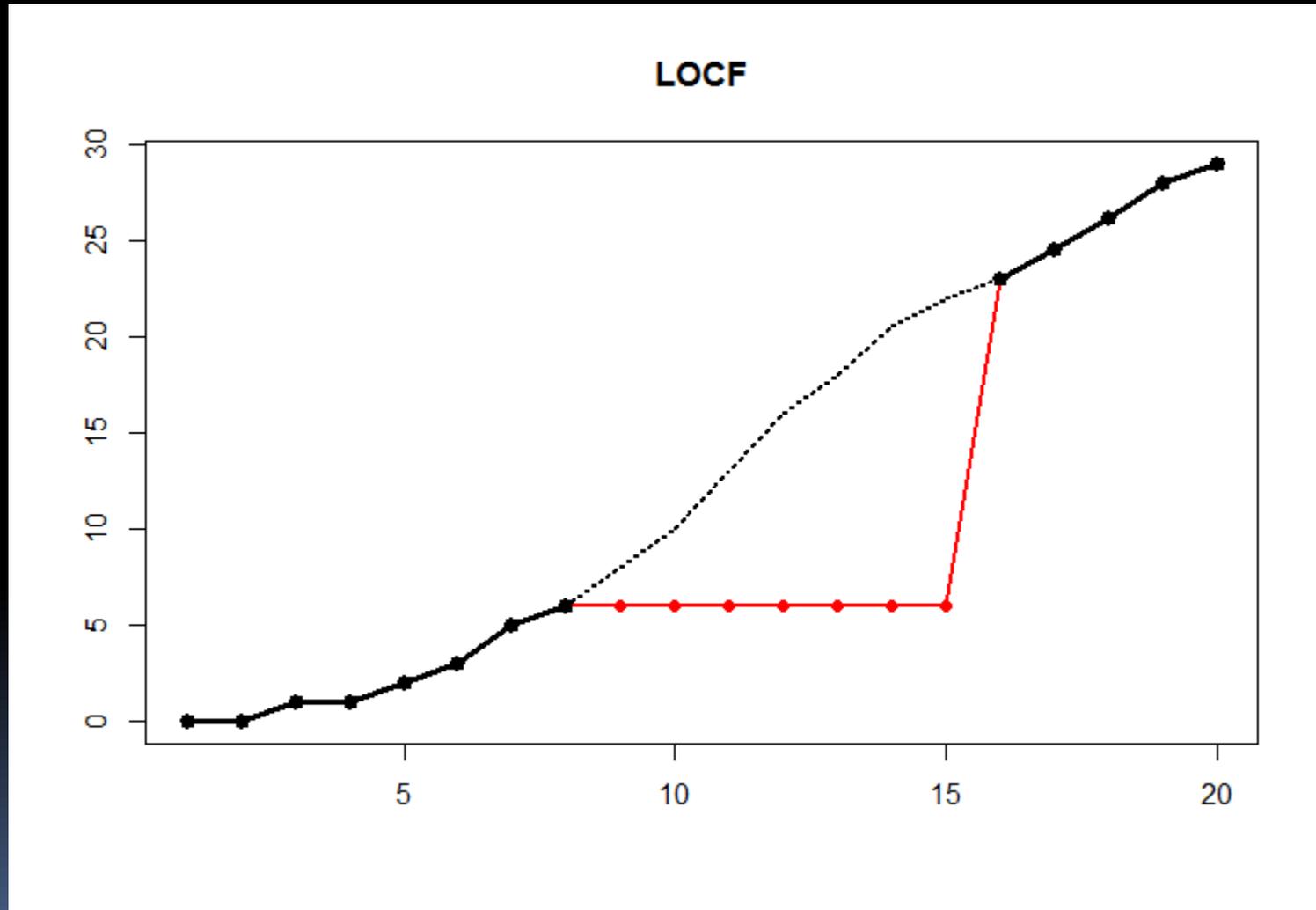
Mais... (LOCF)



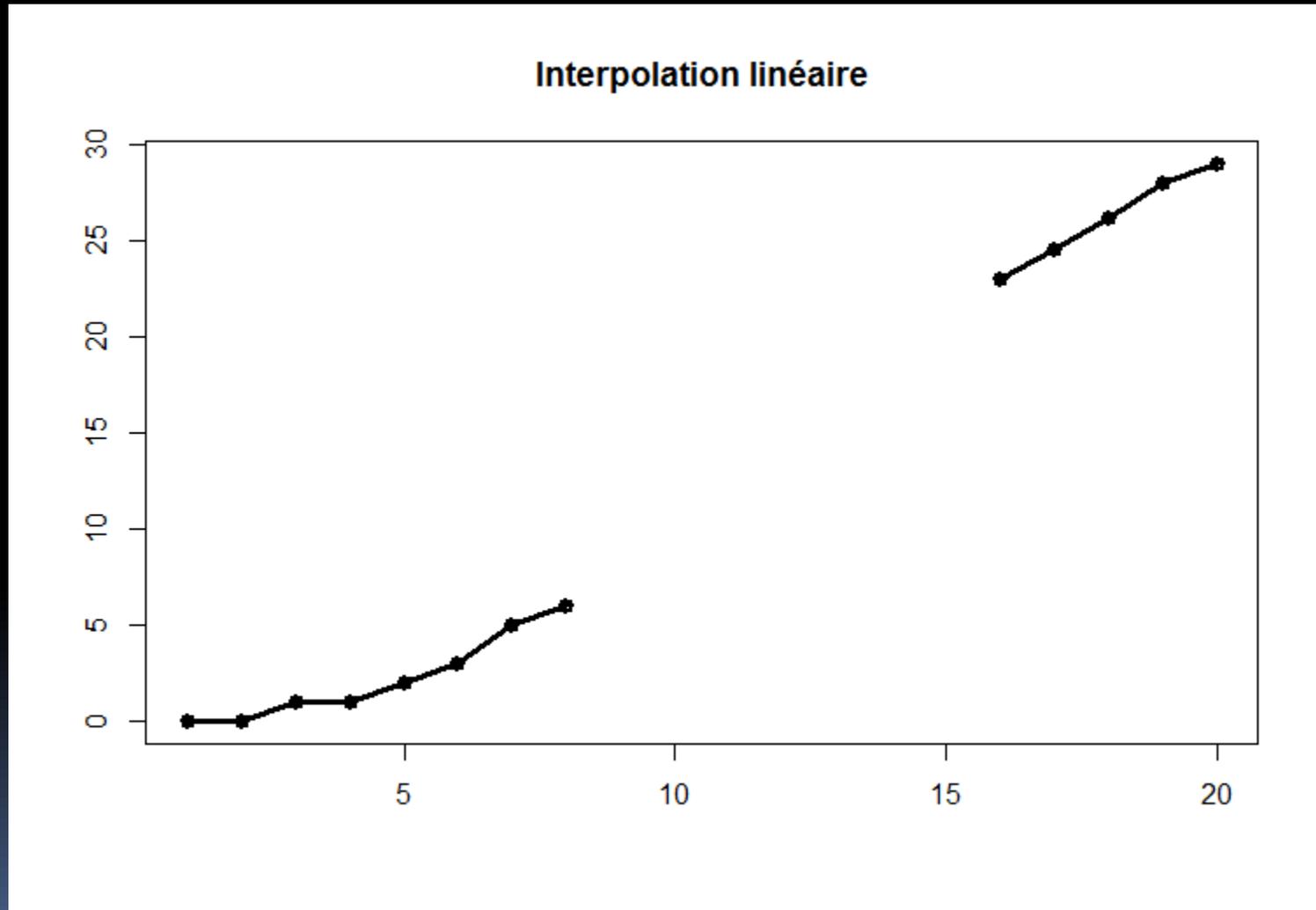
Mais... (LOCF)



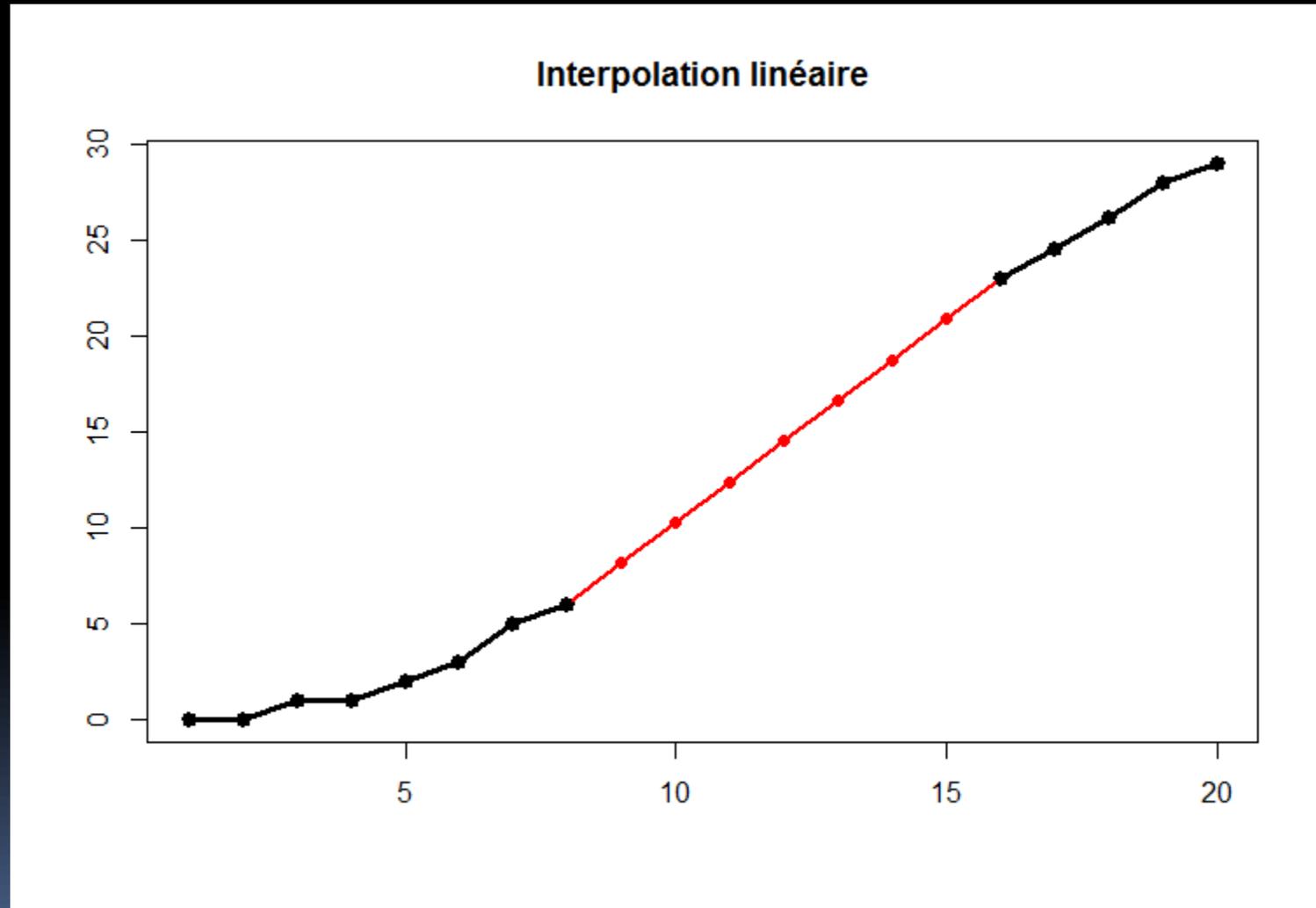
Mais... (LOCF)



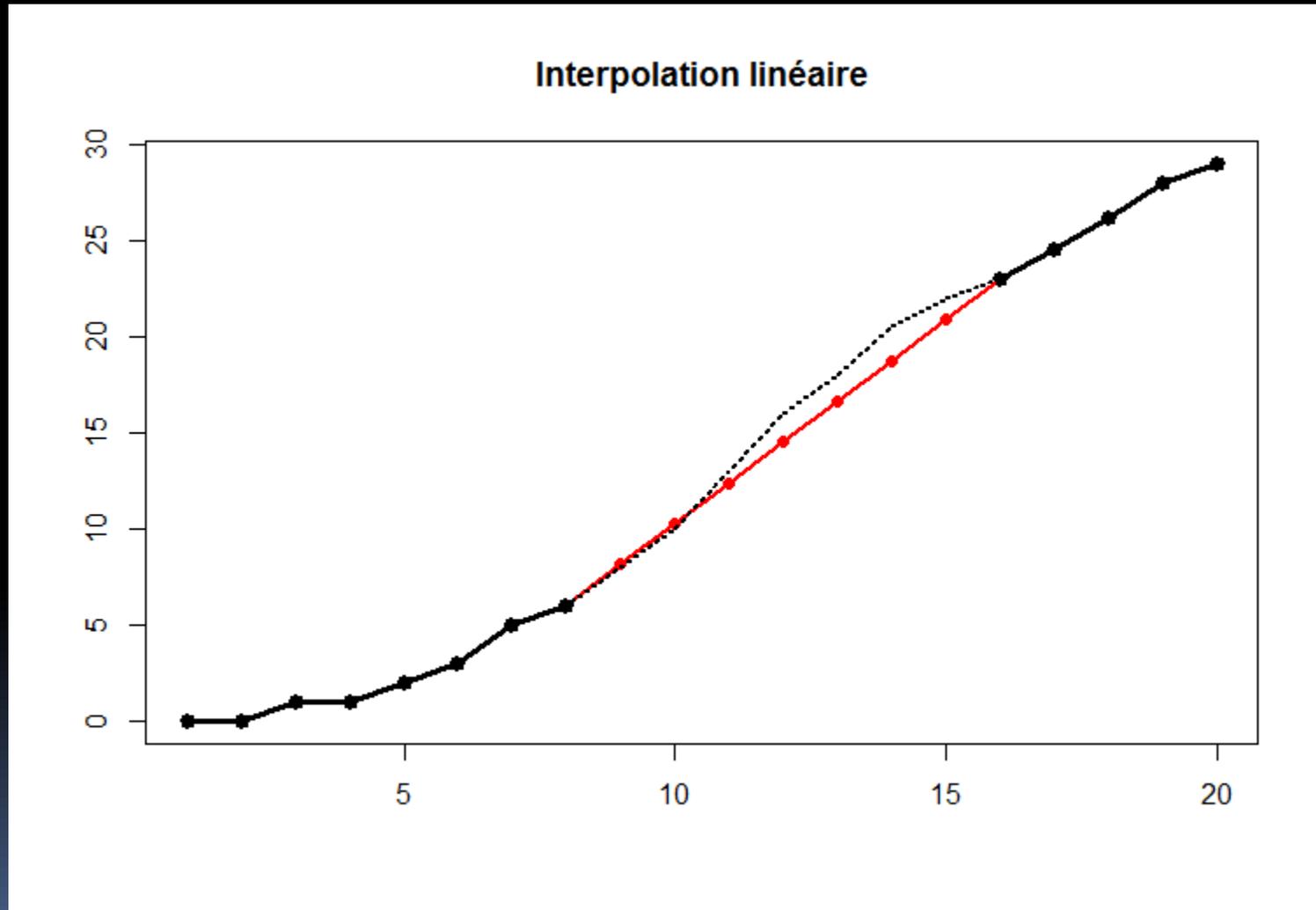
Interpolation linéaire



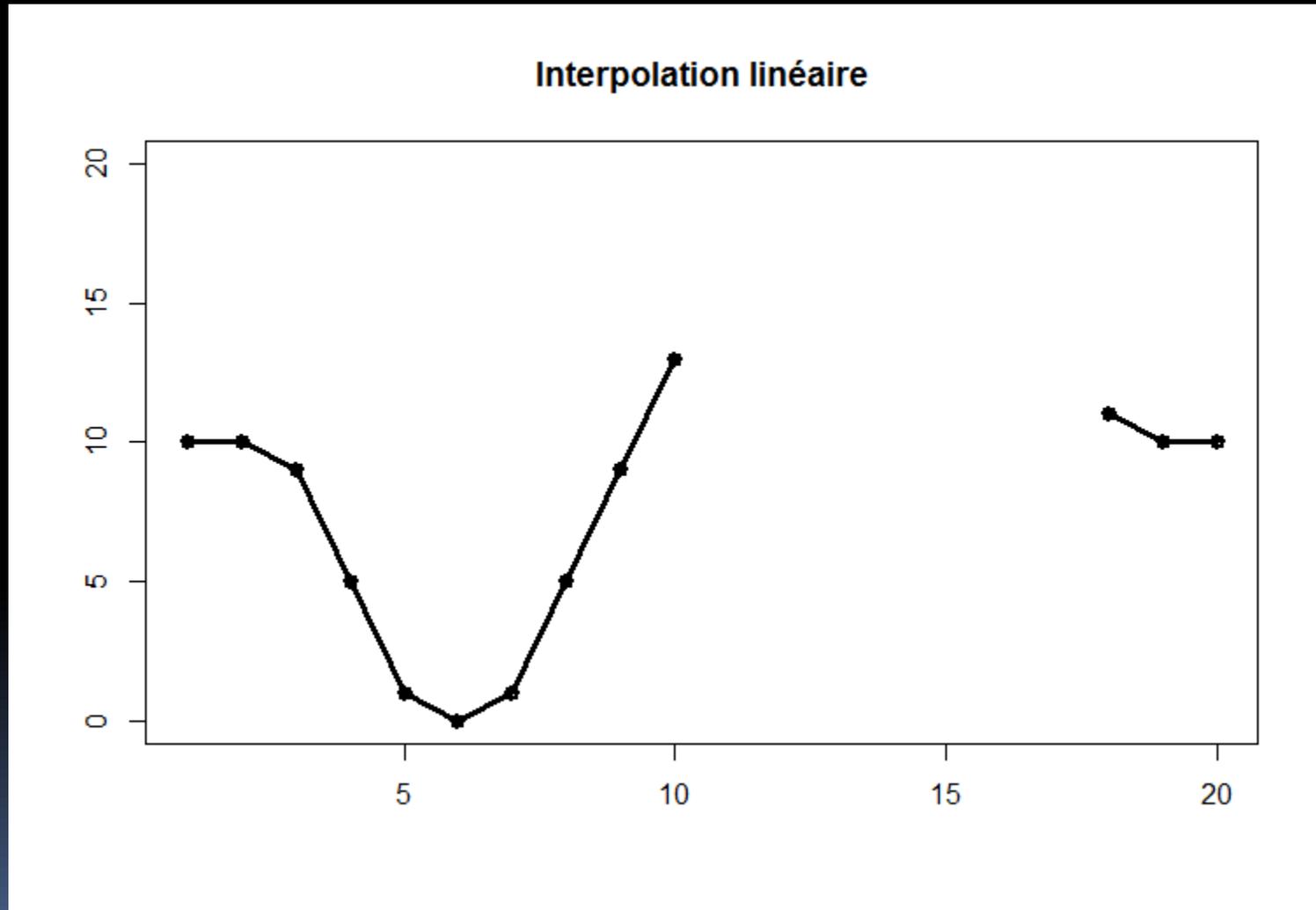
Interpolation linéaire



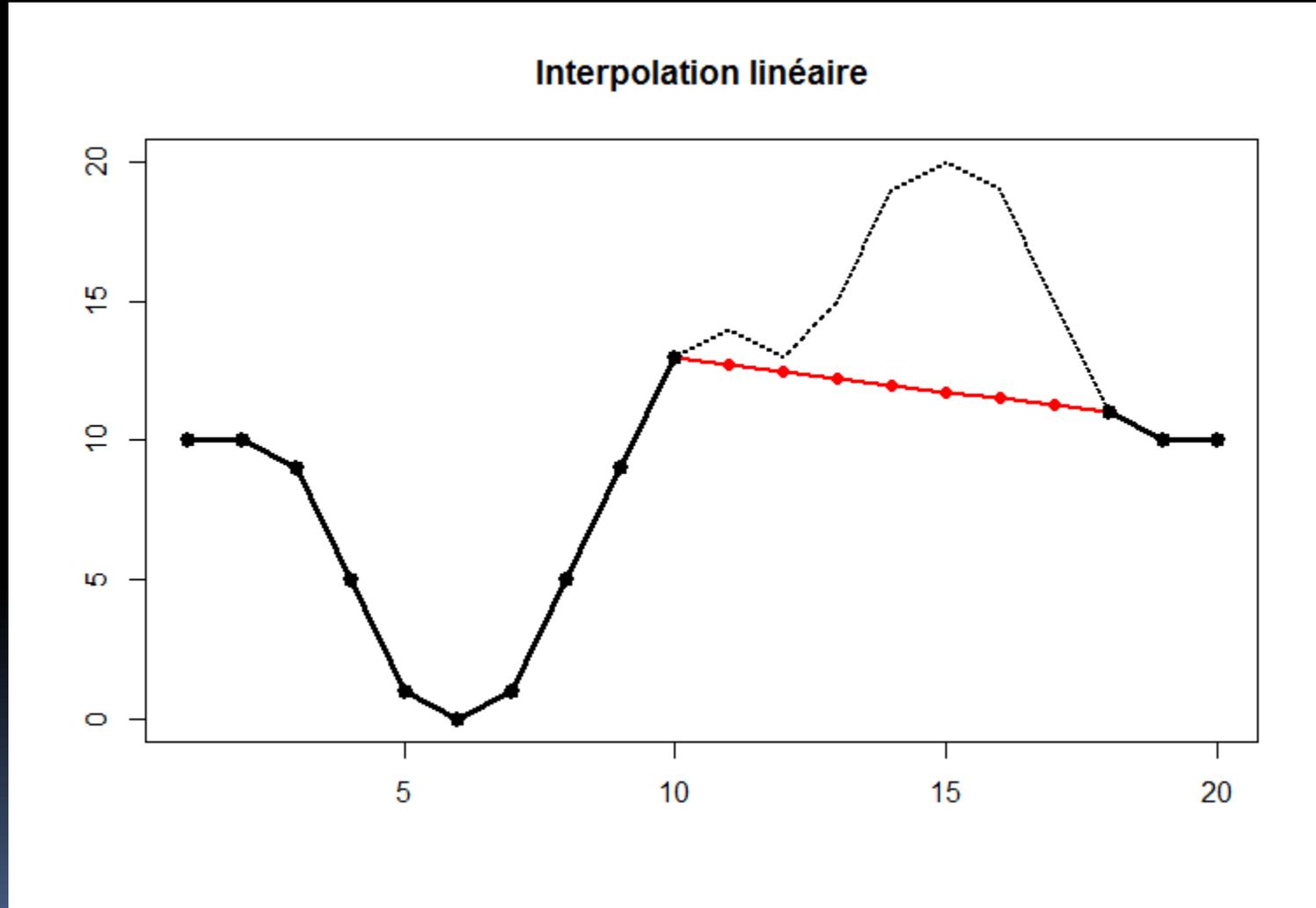
Interpolation linéaire

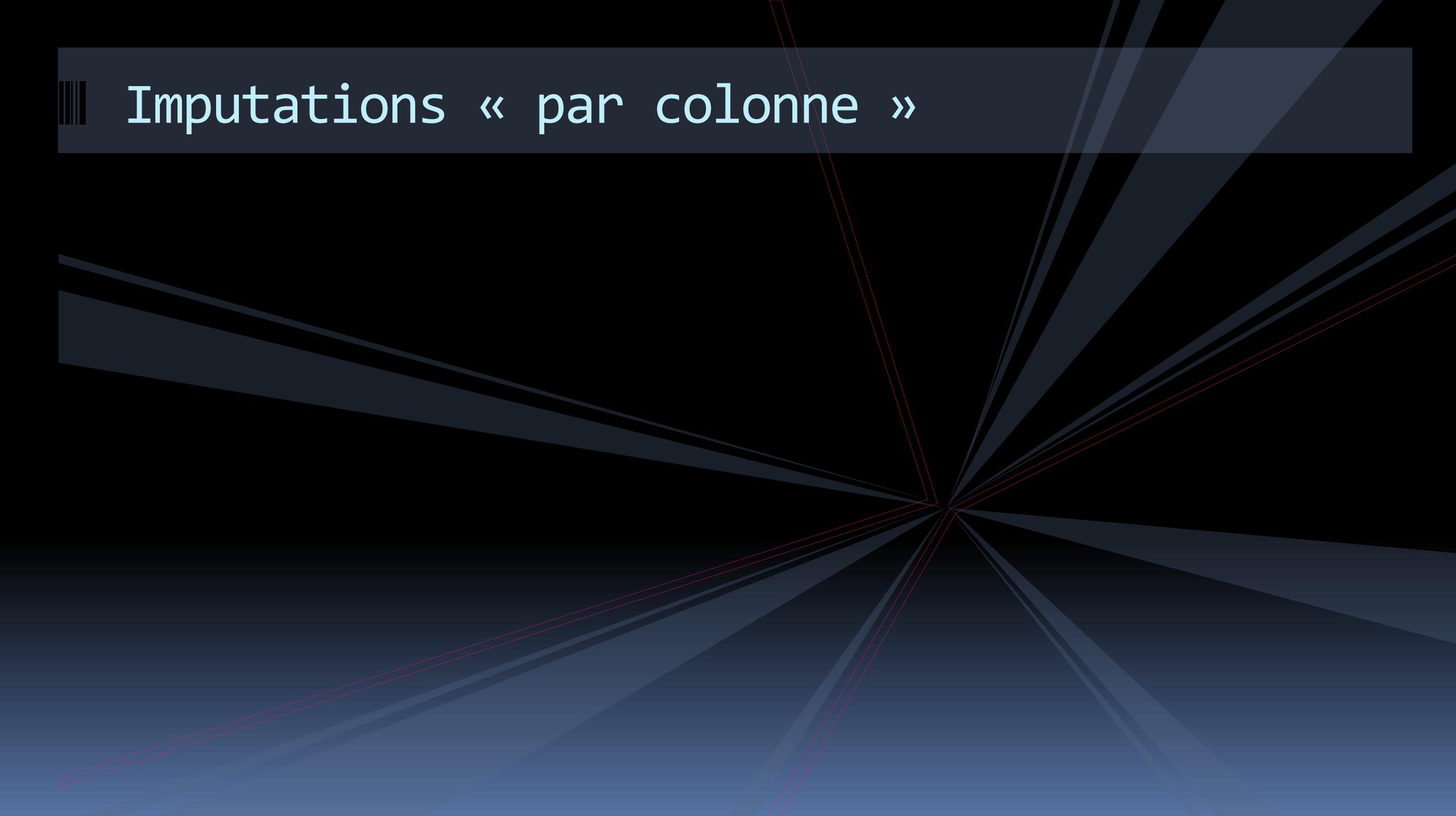


Mais (Interpolation linéaire)



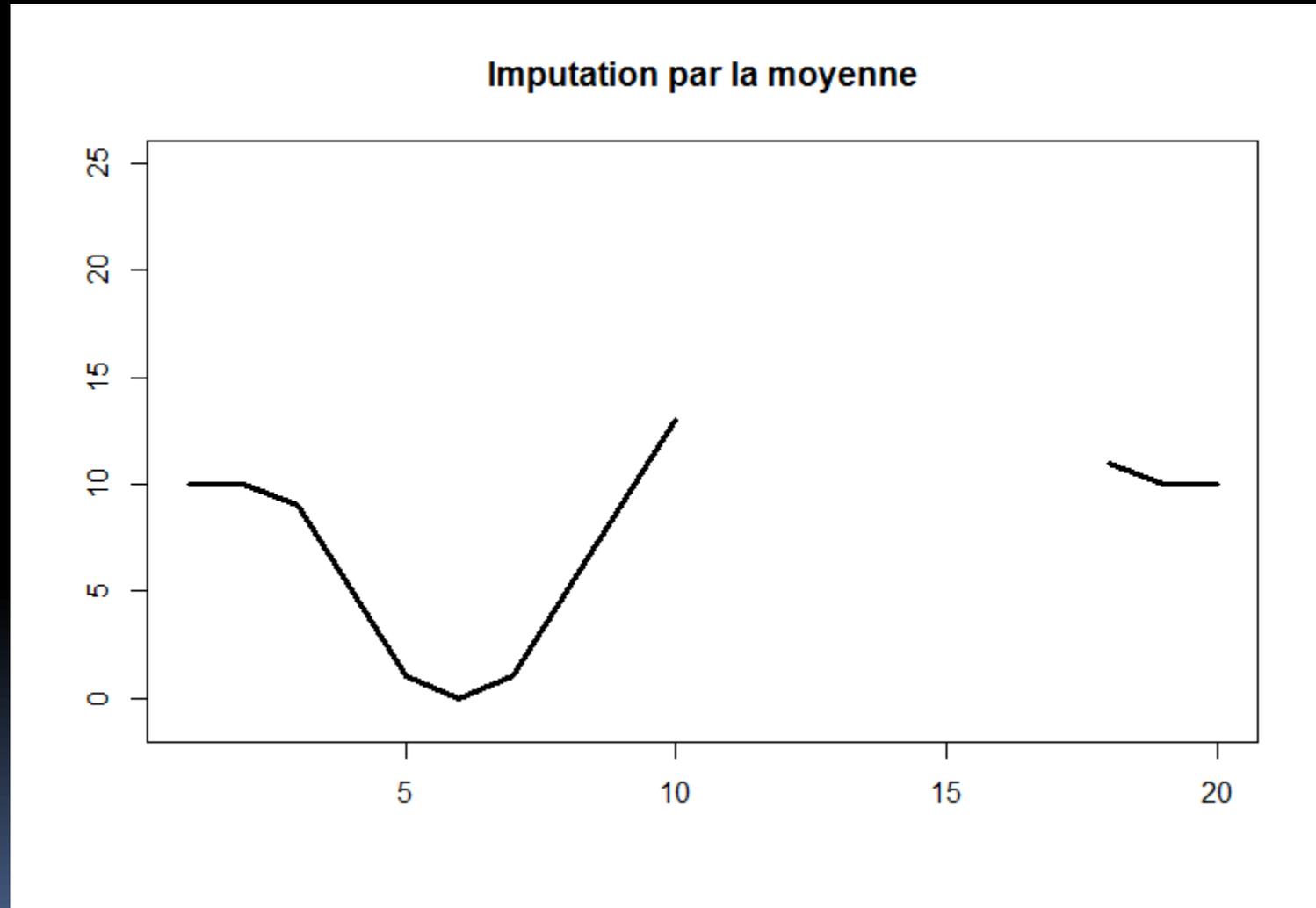
Mais (Interpolation linéaire)



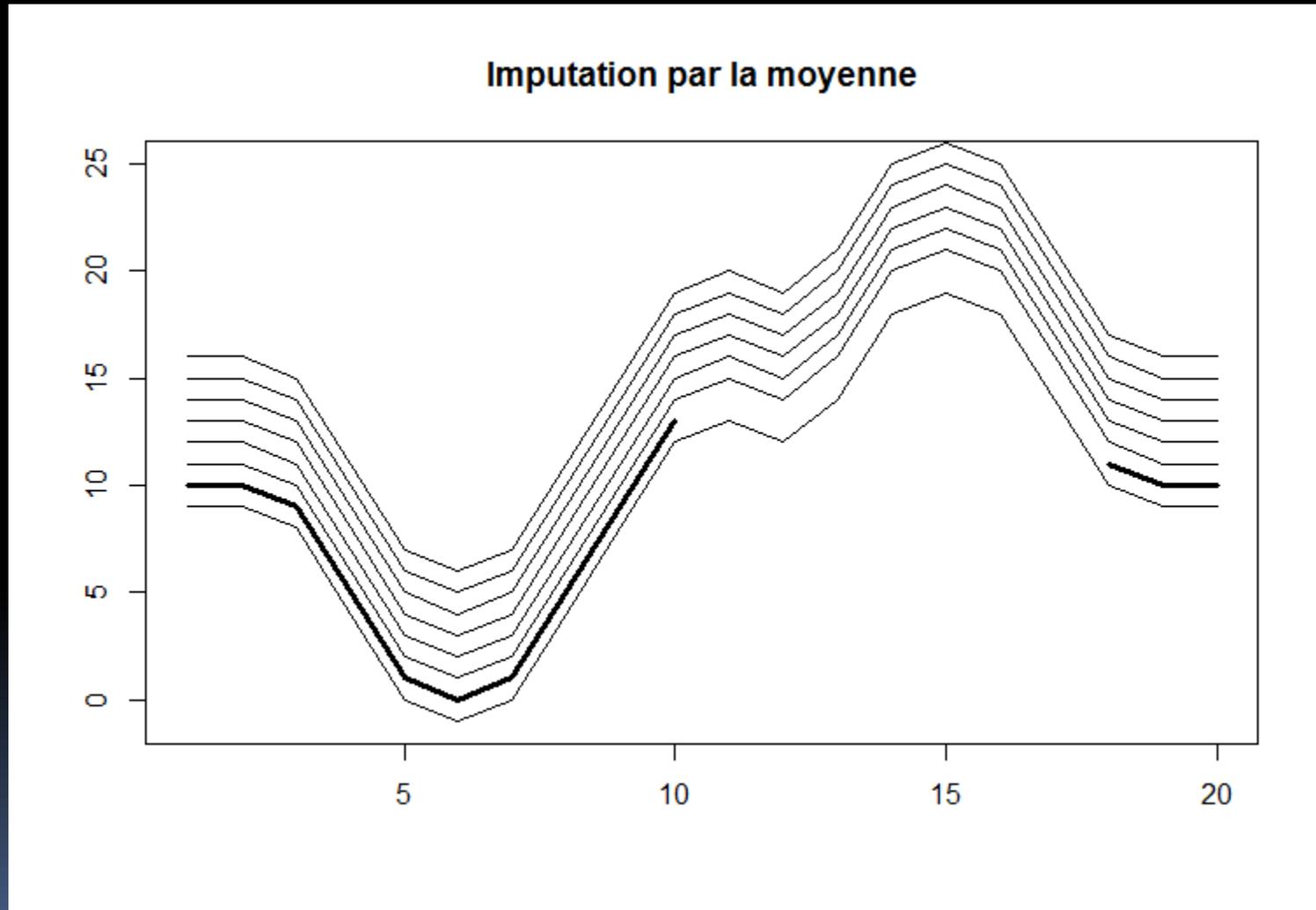


Imputations « par colonne »

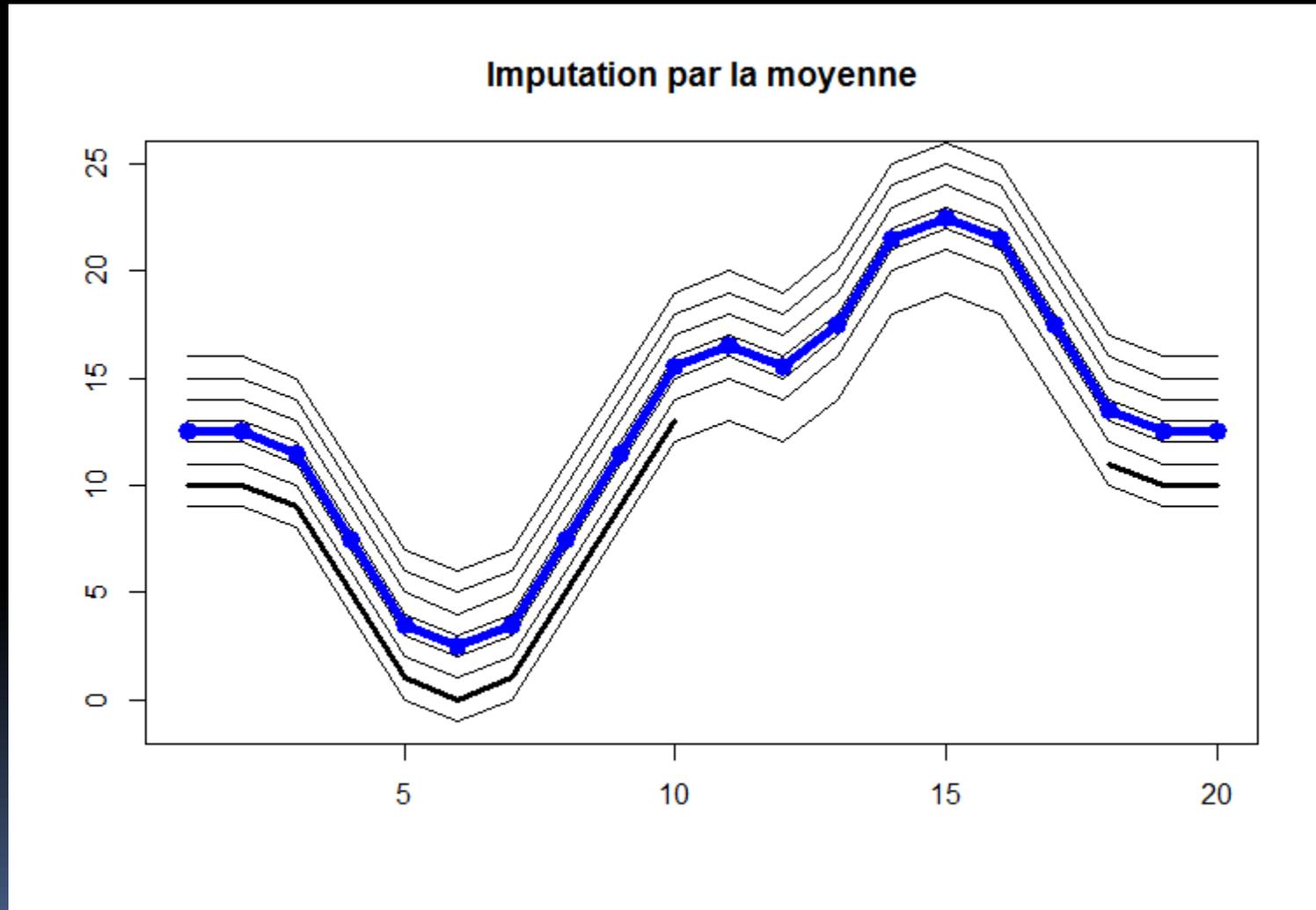
Imputation par la moyenne



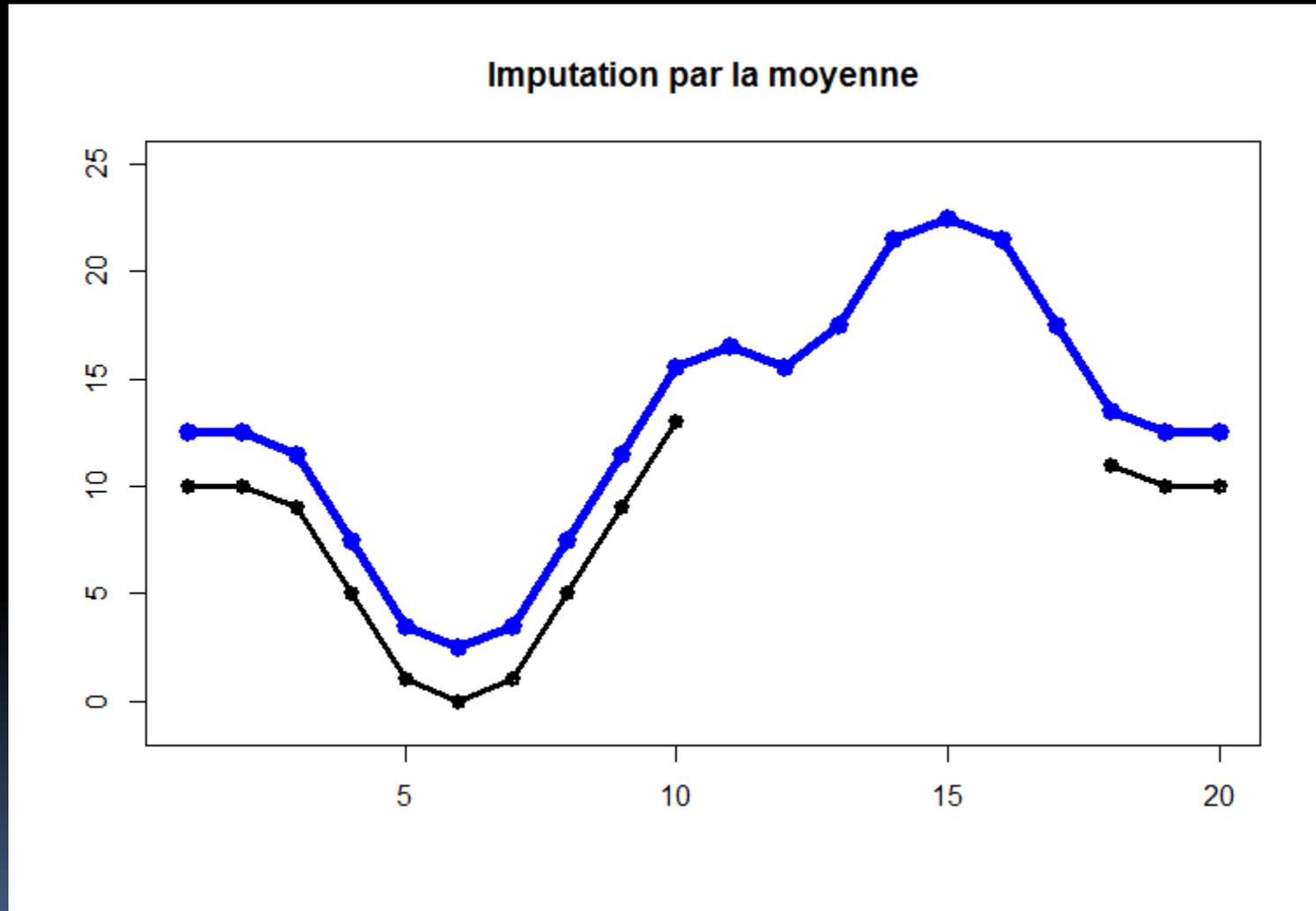
Imputation par la moyenne



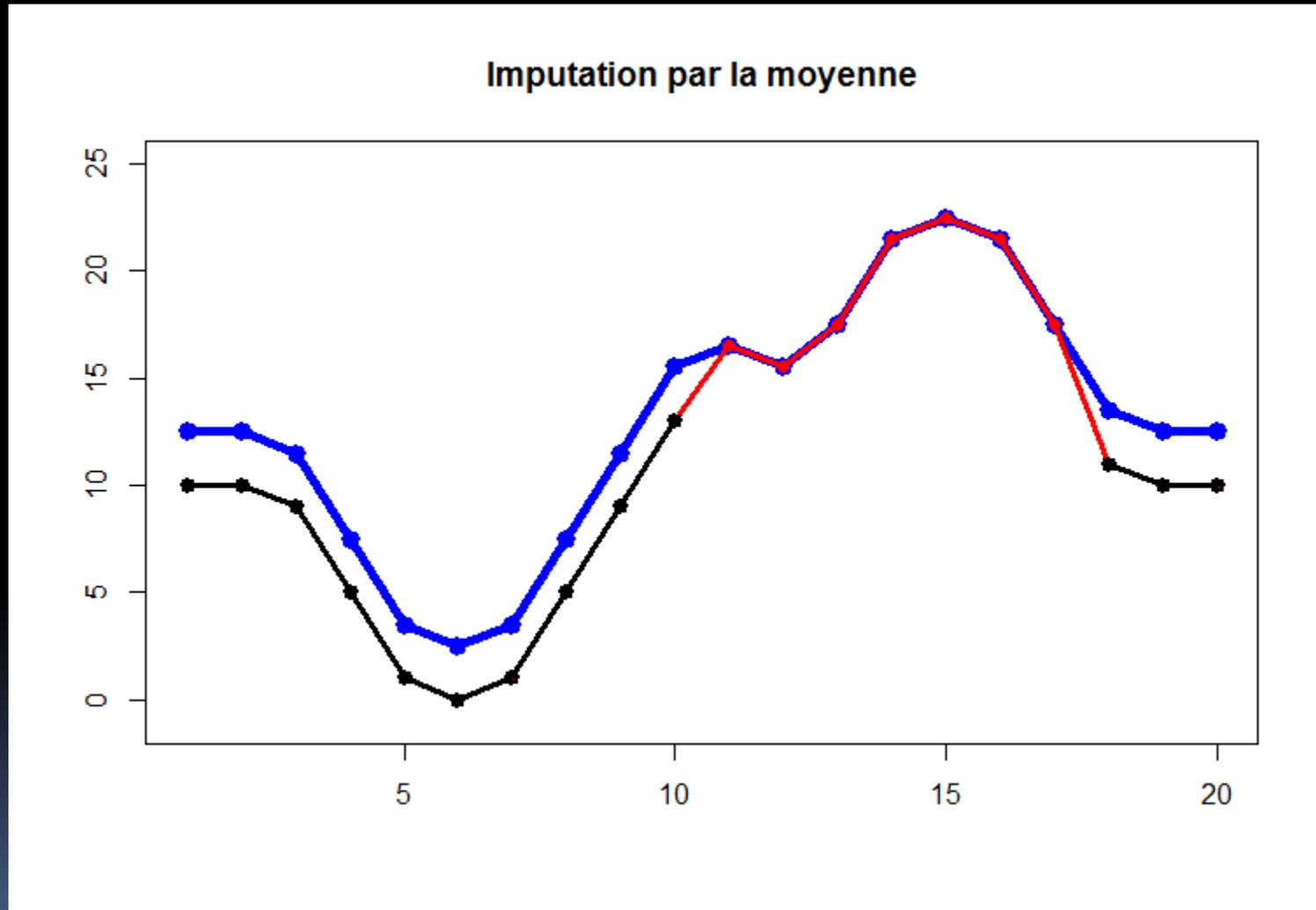
Imputation par la moyenne



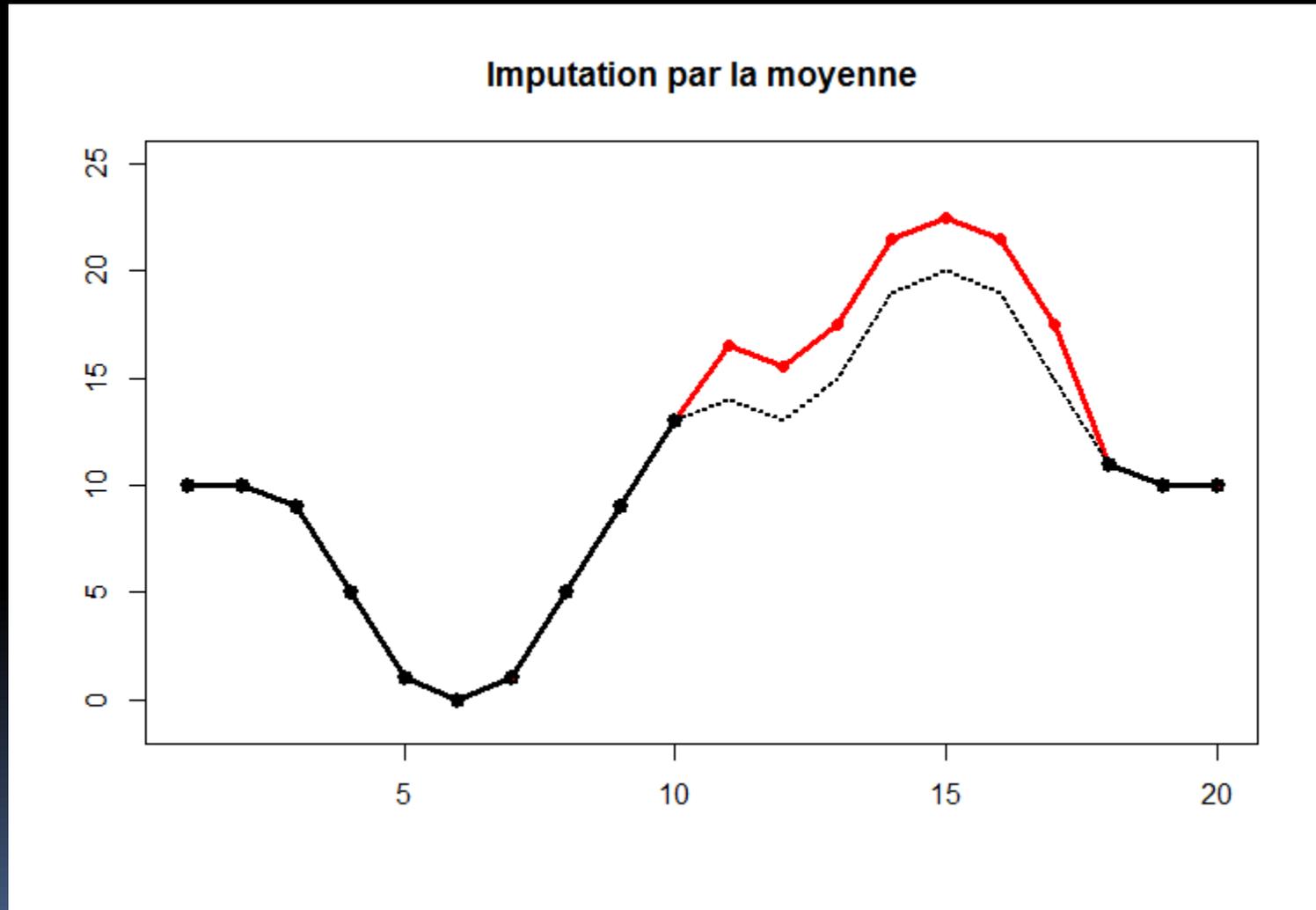
Imputation par la moyenne



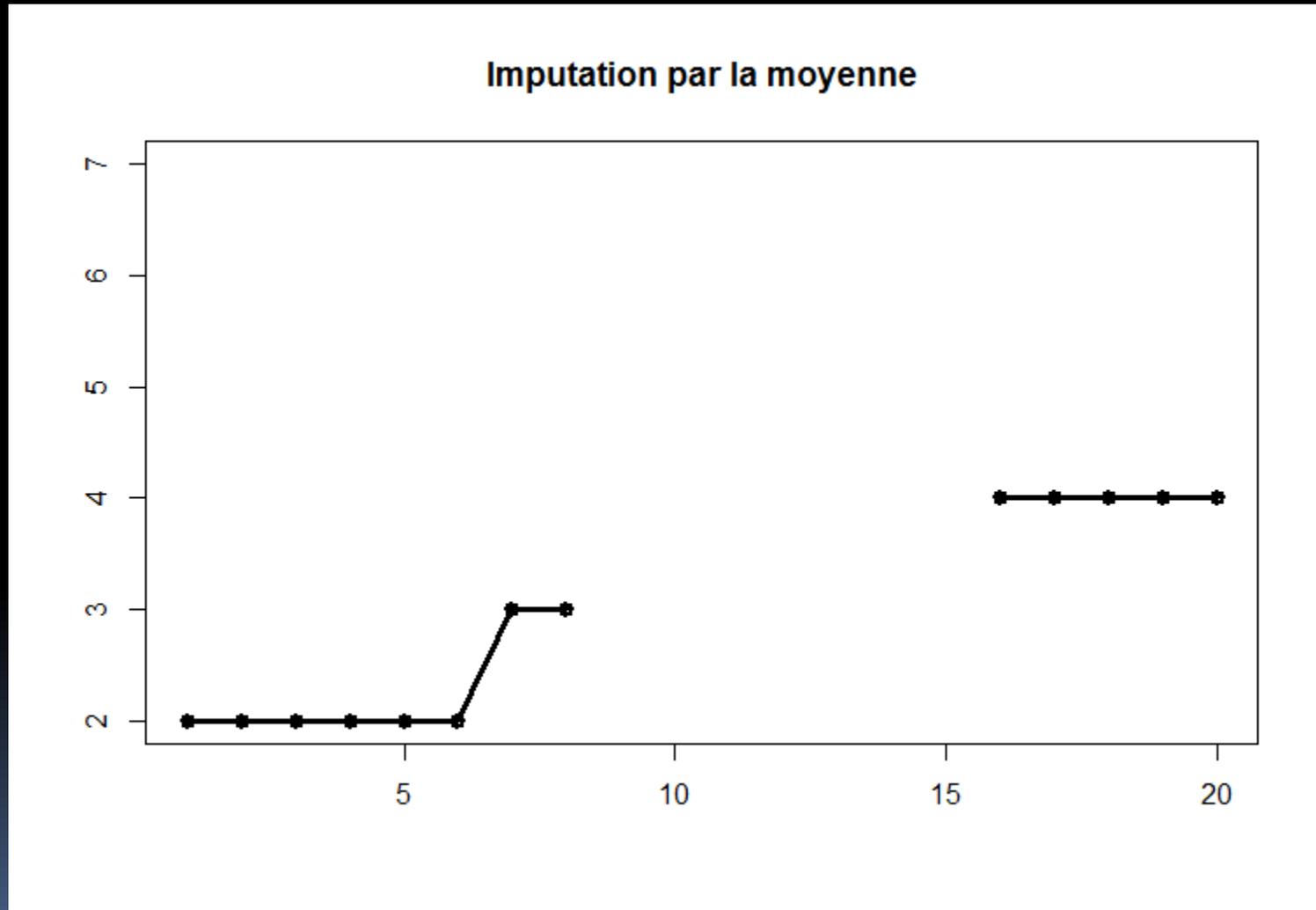
Imputation par la moyenne



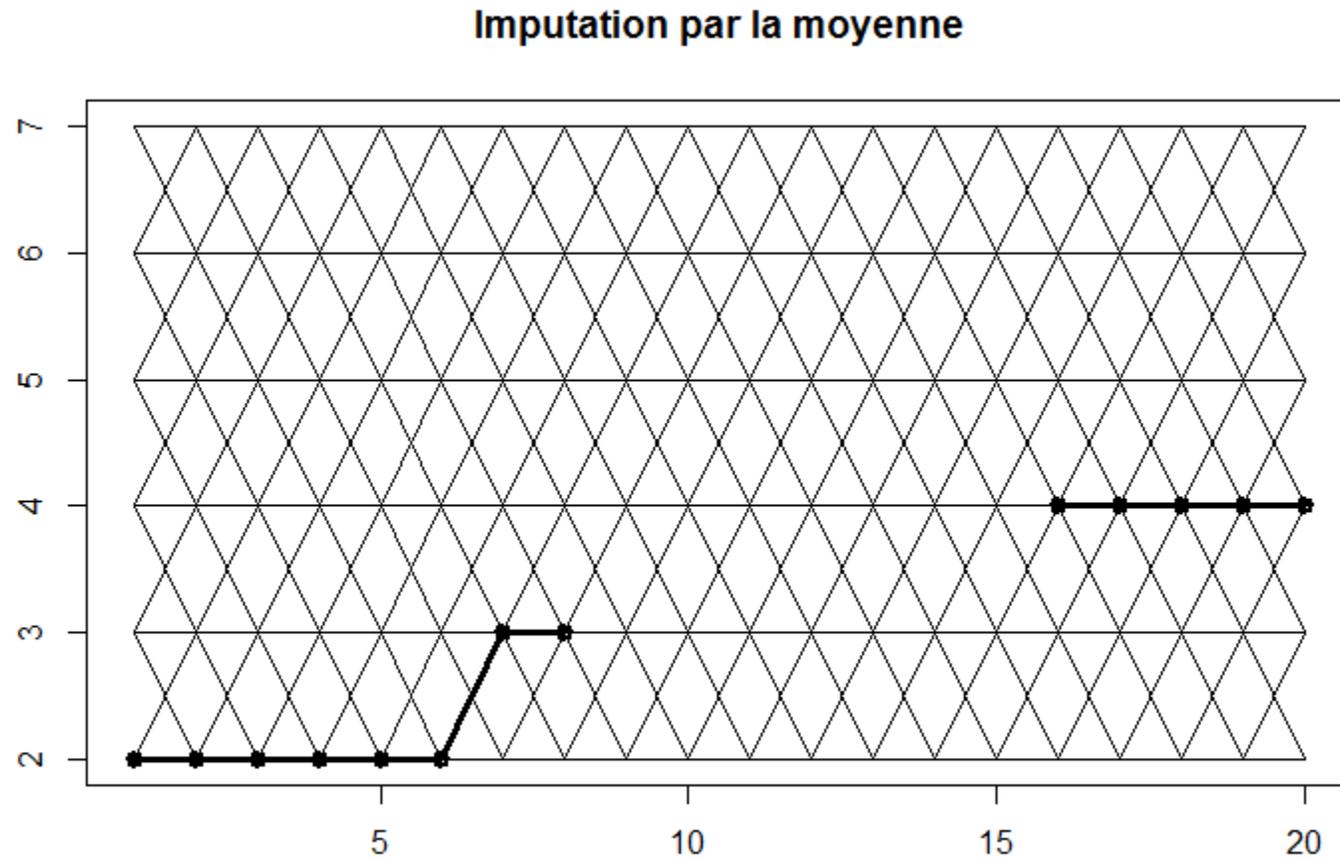
Imputation par la moyenne



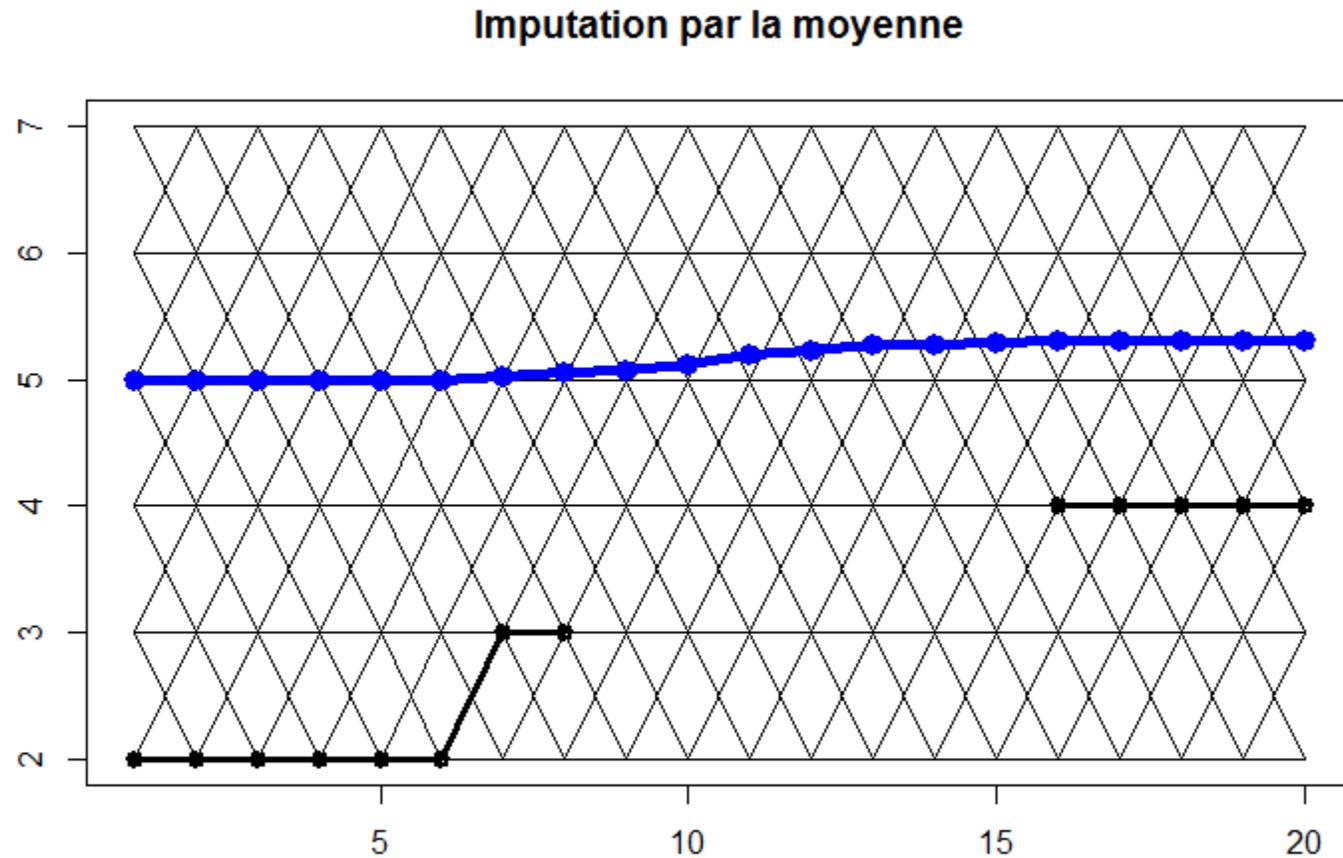
Mais... (Imputation par la moyenne)



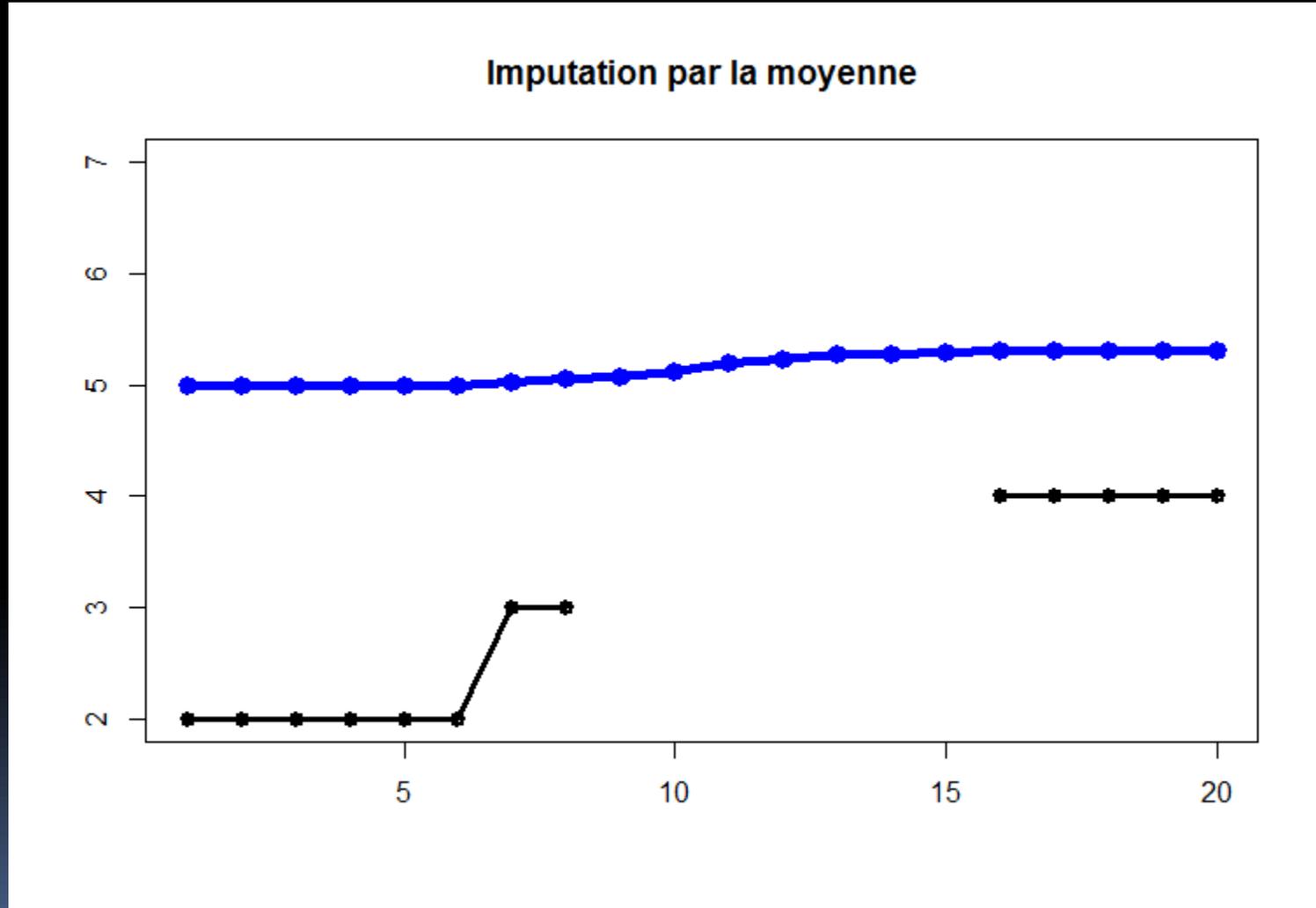
Mais... (Imputation par la moyenne)



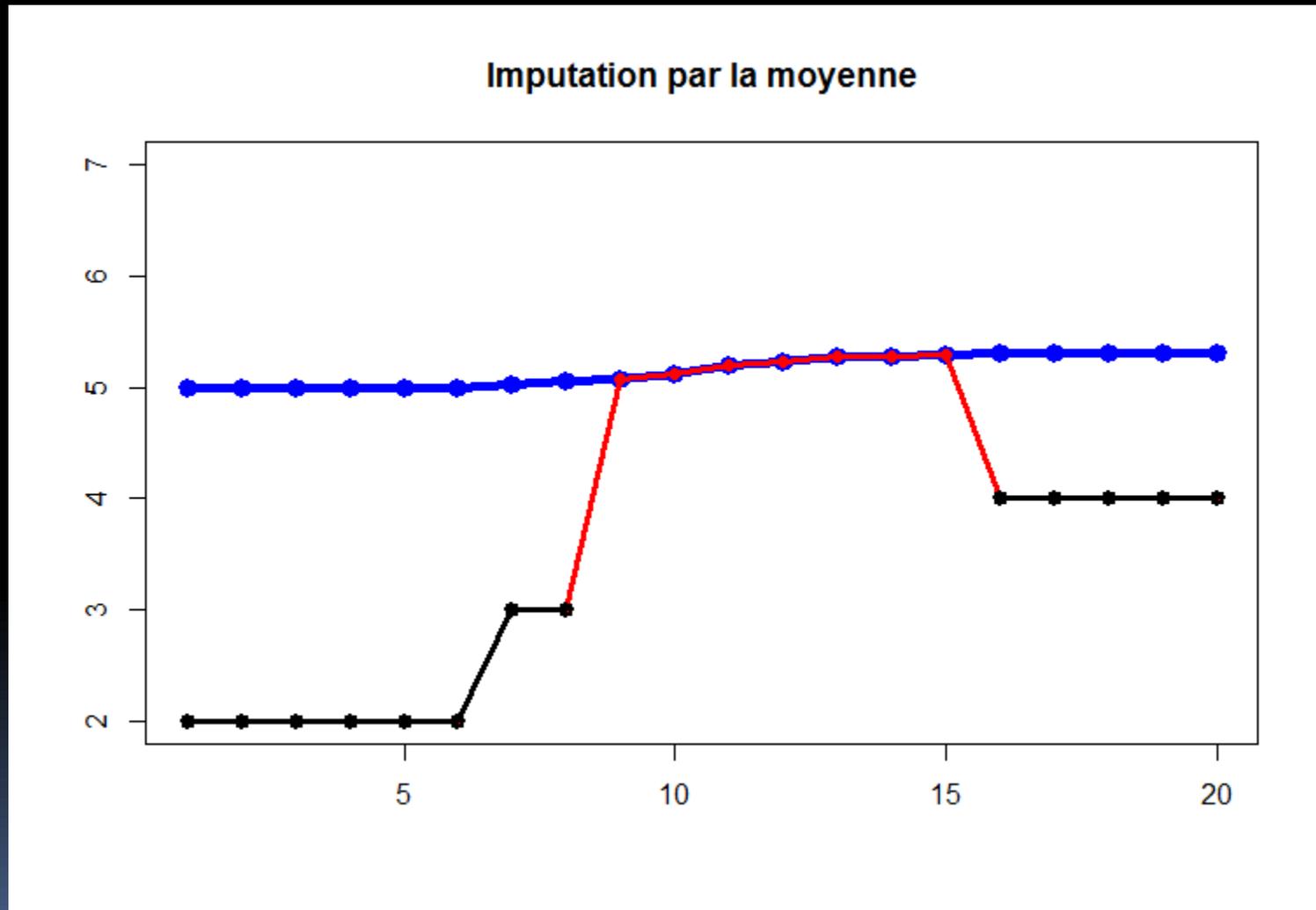
Mais... (Imputation par la moyenne)



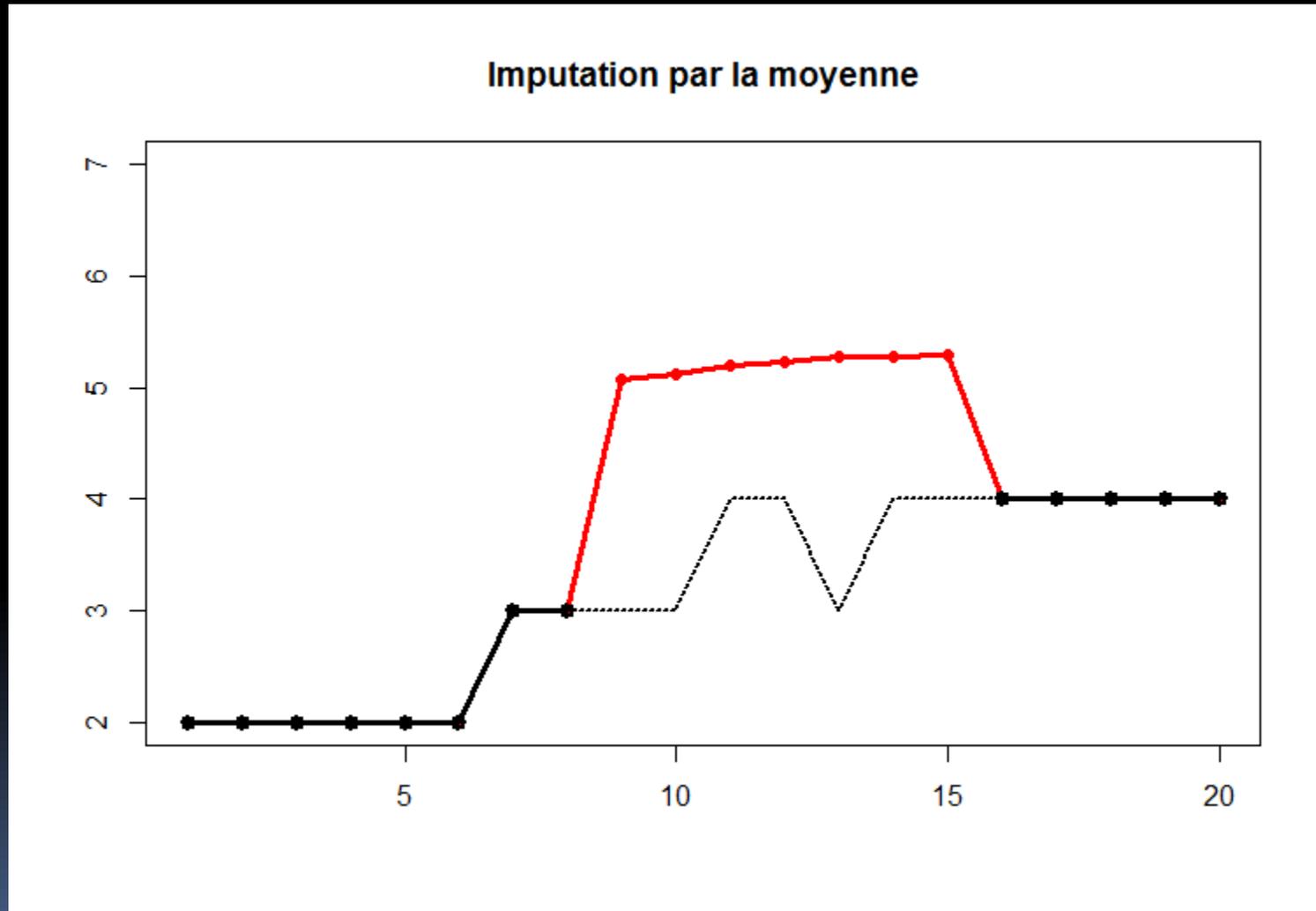
Mais... (Imputation par la moyenne)



Mais... (Imputation par la moyenne)



Mais... (Imputation par la moyenne)



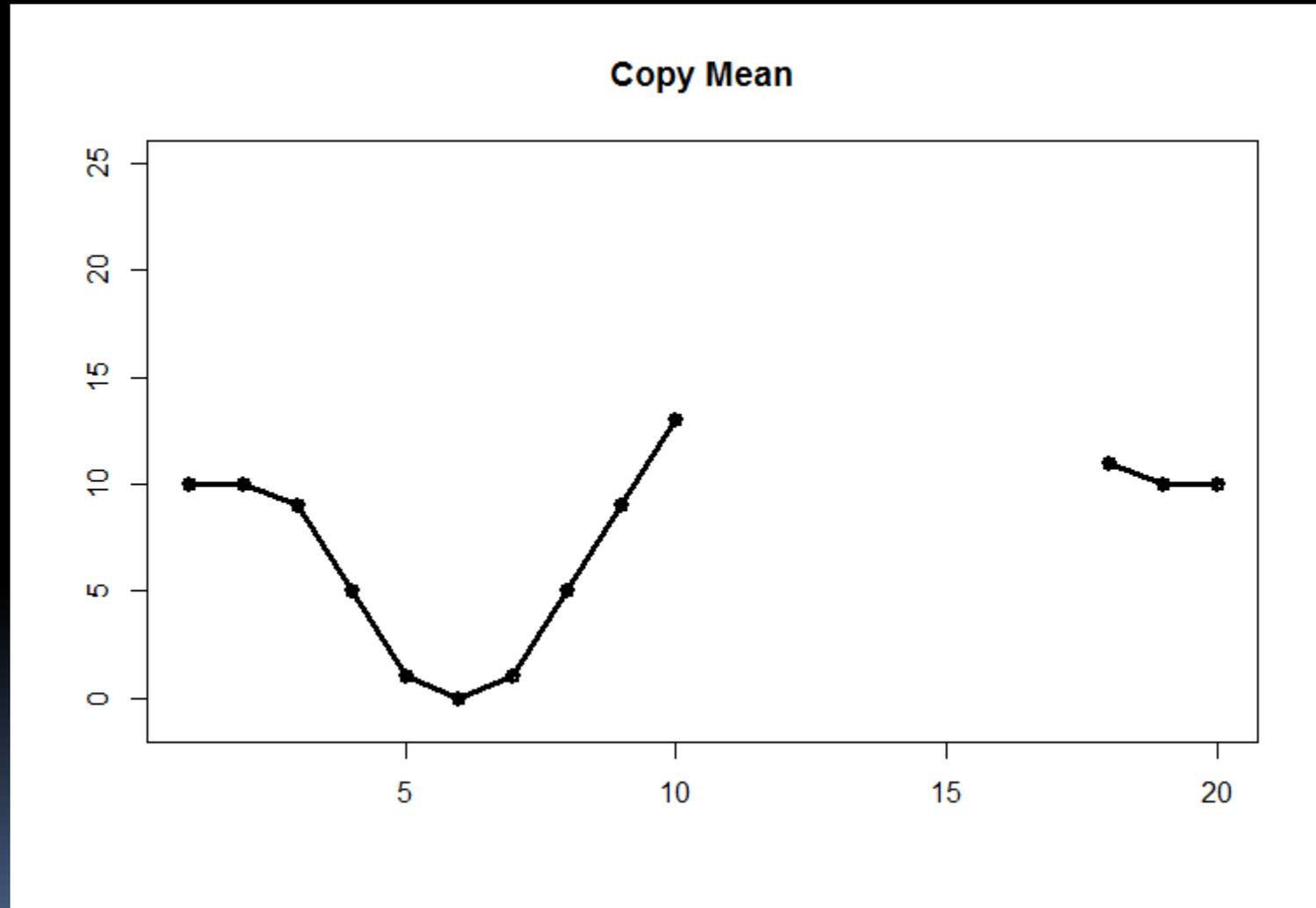
Bilan (pas brillant !)

- Imputation « par trajectoires »
 - + Prend en compte la spécificité individuelle
 - Insensibles aux évènements locaux
- Imputation « par colonnes »
 - Ne prend pas en compte la spécificité individuelle
 - + Considère les évènements locaux

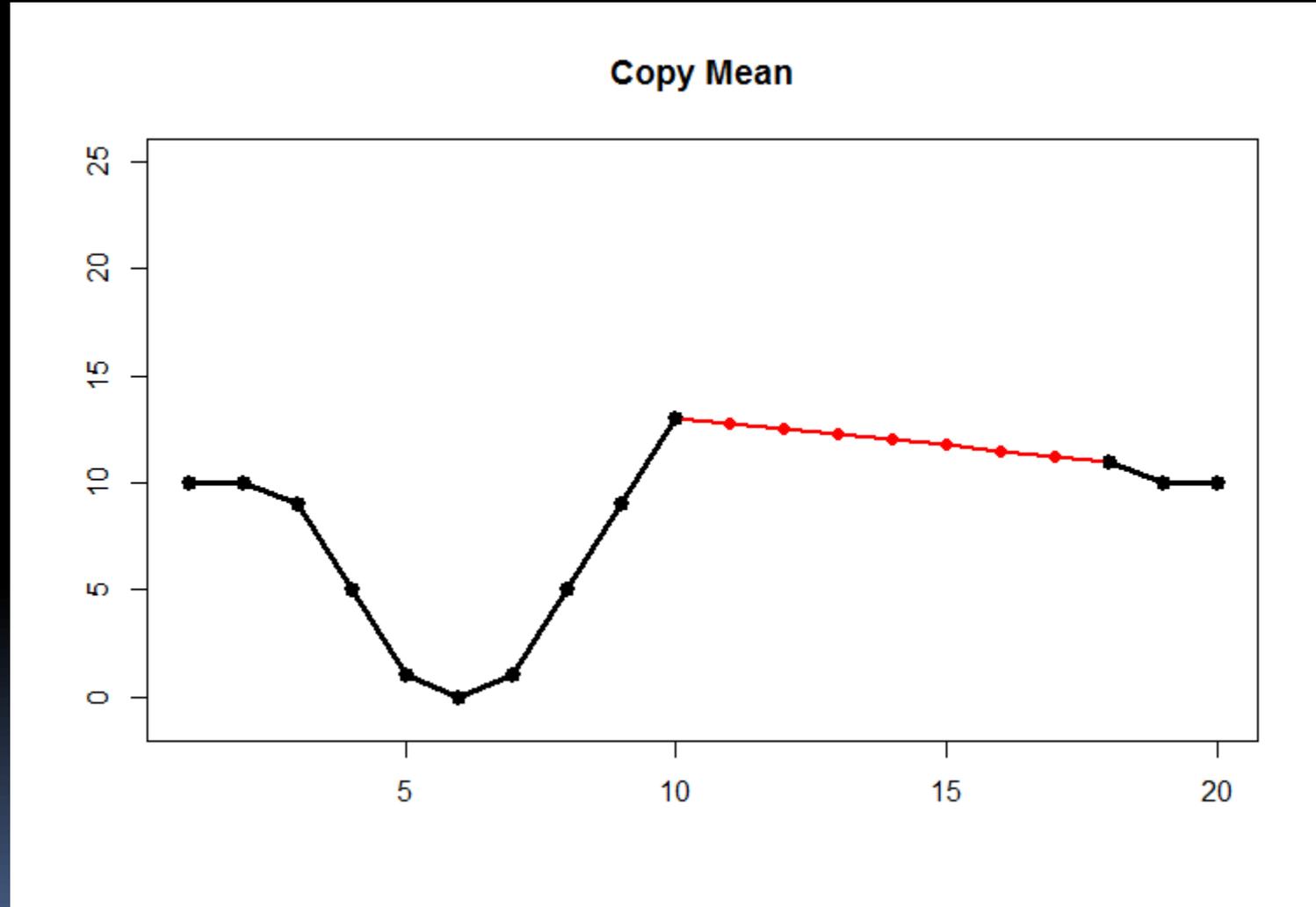


copyMean : imputation par trajectoire et
par colonne

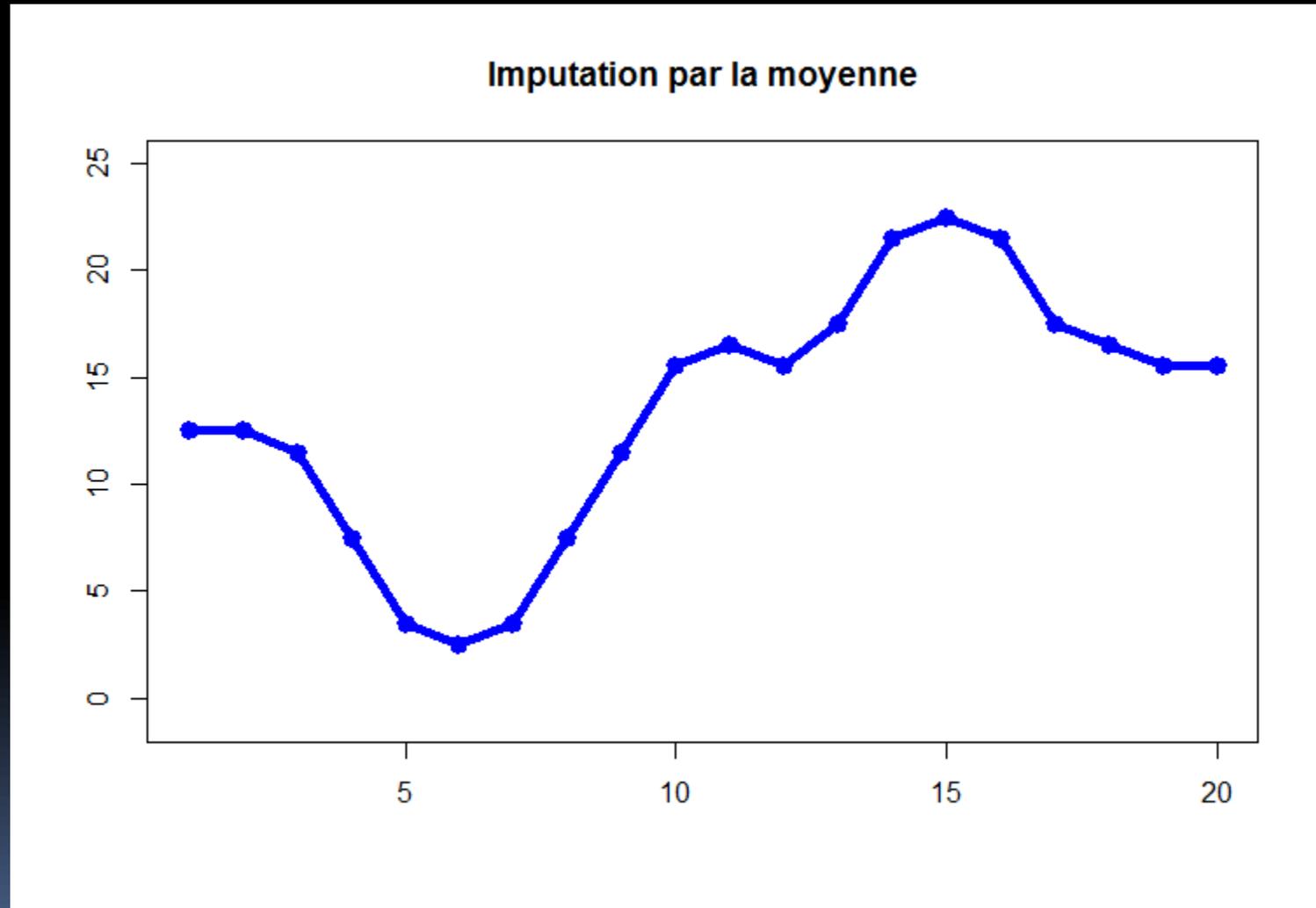
Phase I : interpolation linéaire



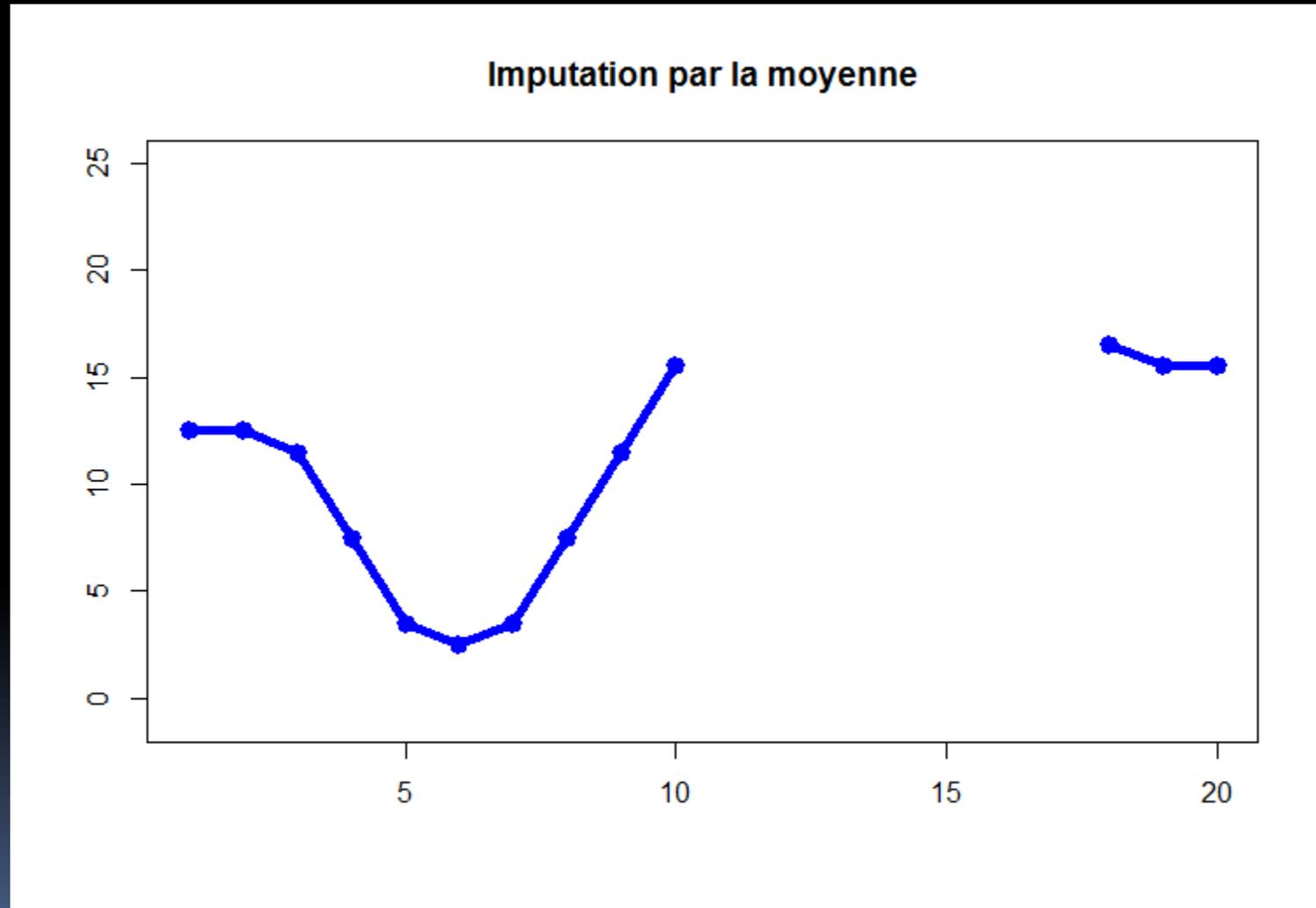
Phase I : interpolation linéaire



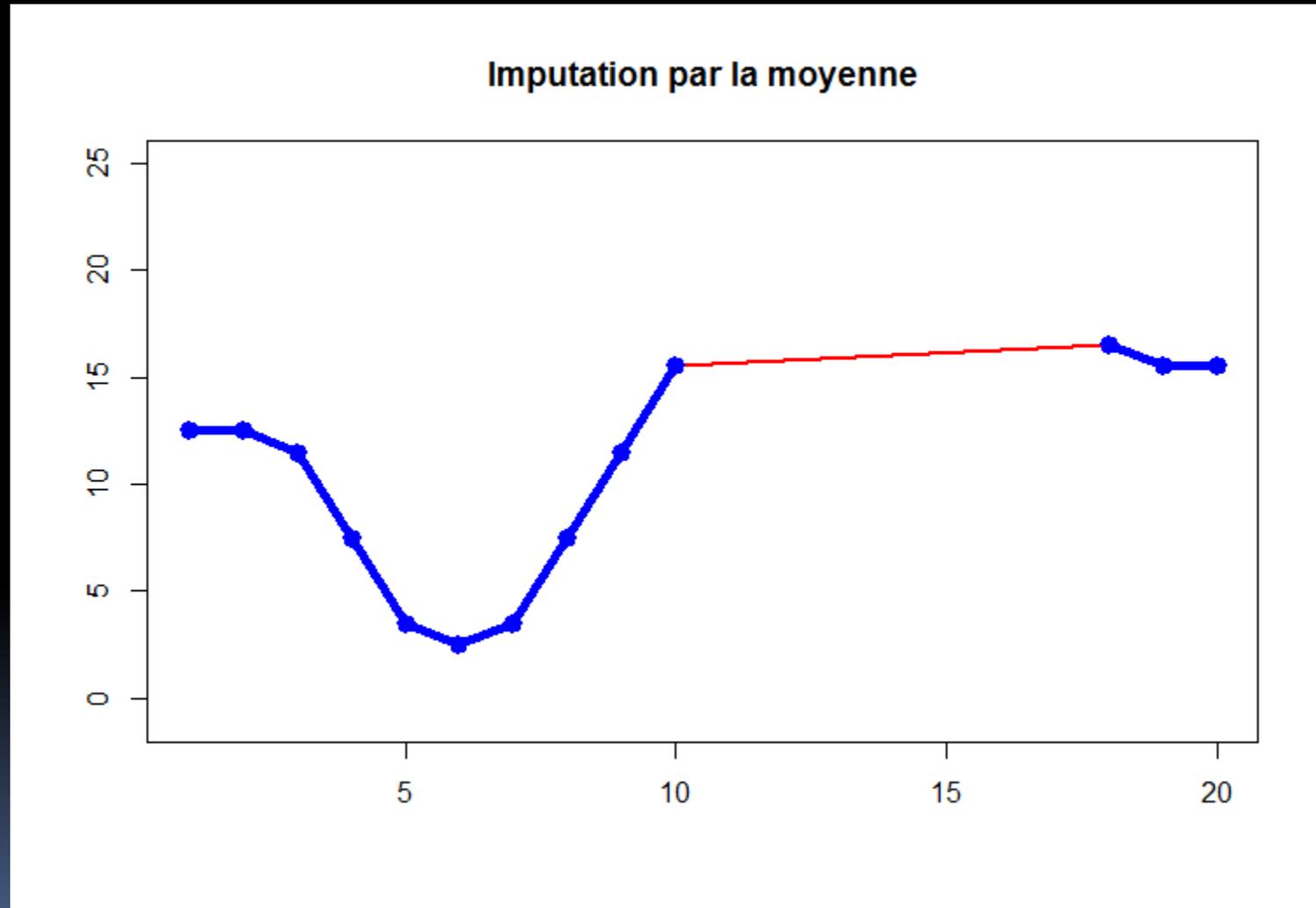
Phase II : ajout d'une variation



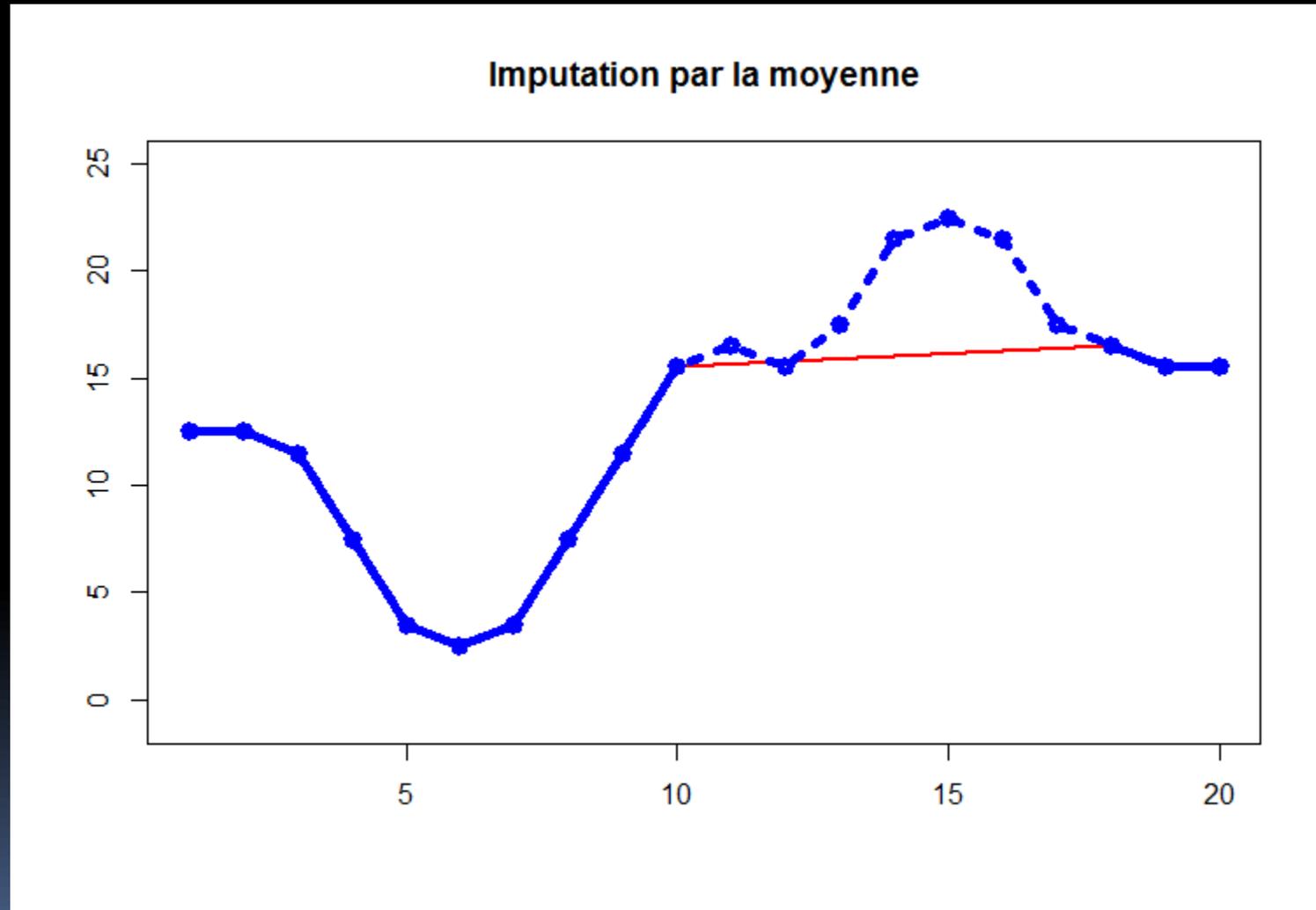
Phase II : ajout d'une variation



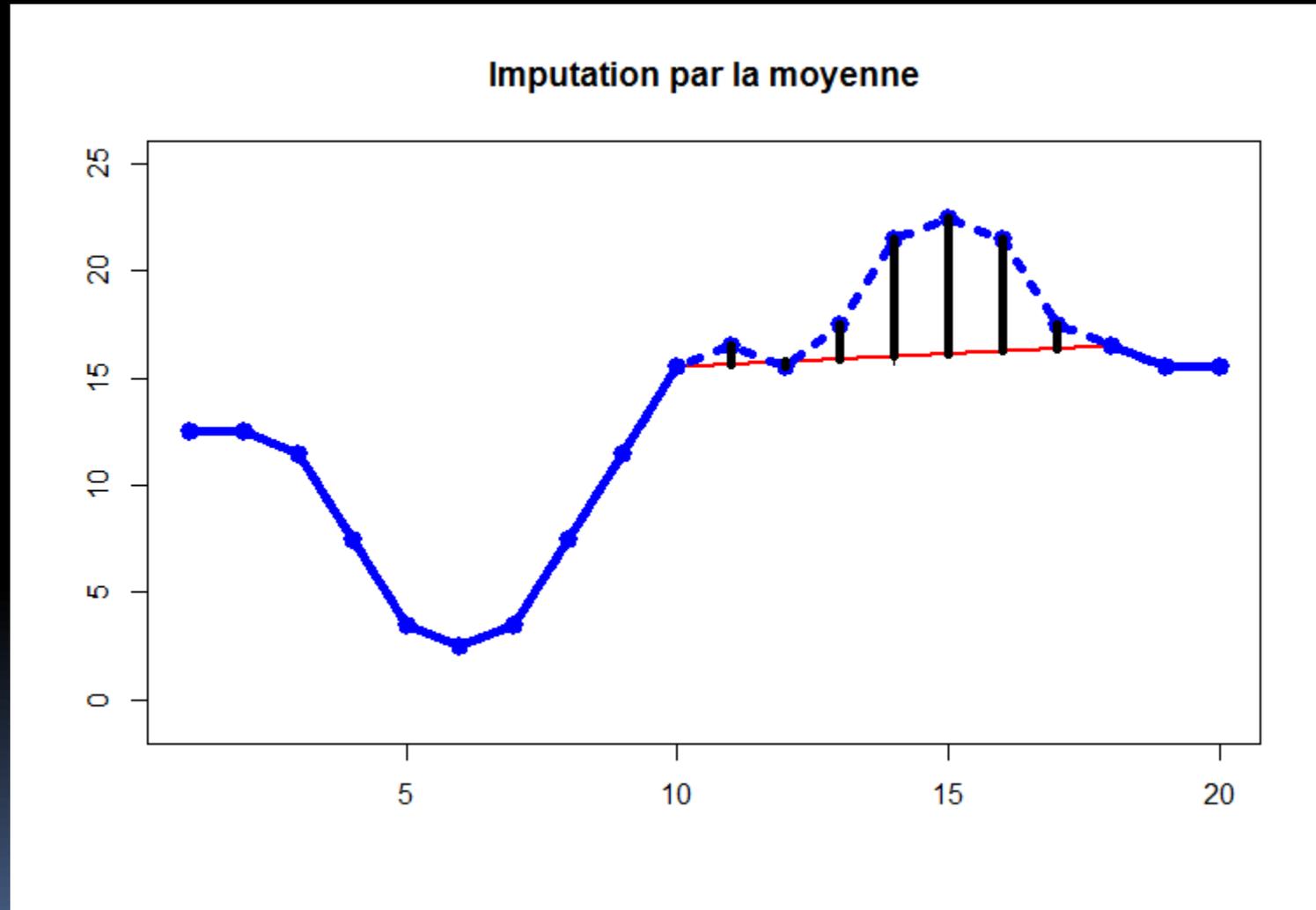
Phase II : ajout d'une variation



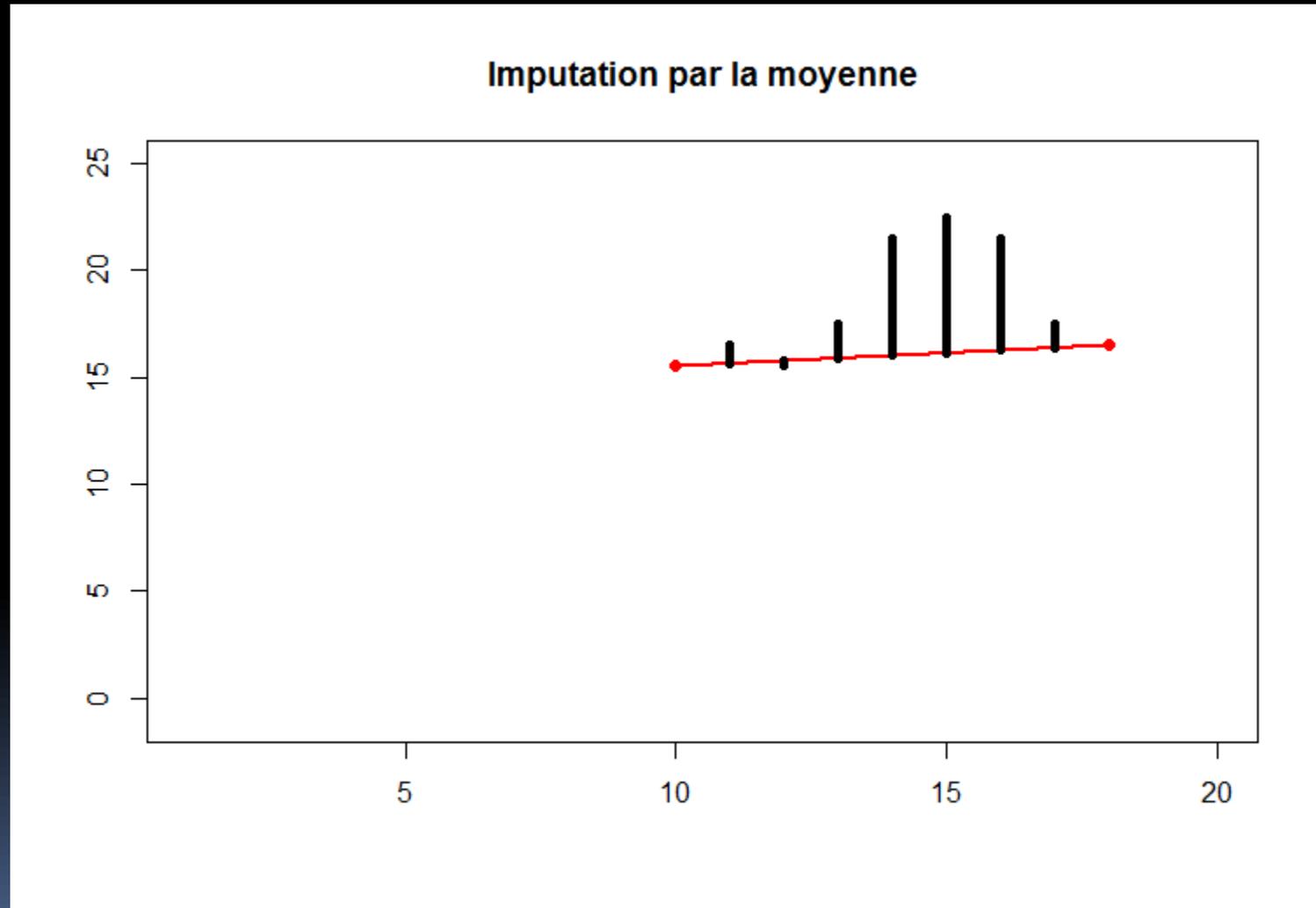
Phase II : ajout d'une variation



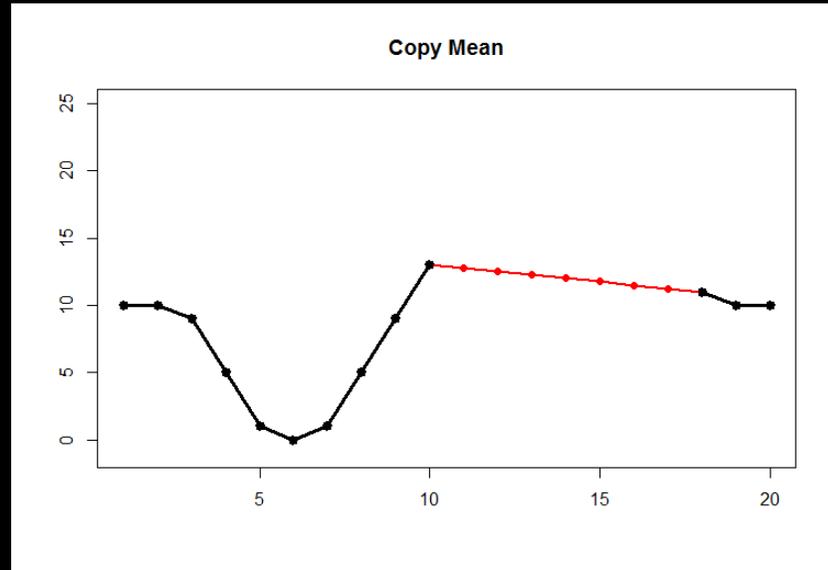
Phase II : ajout d'une variation



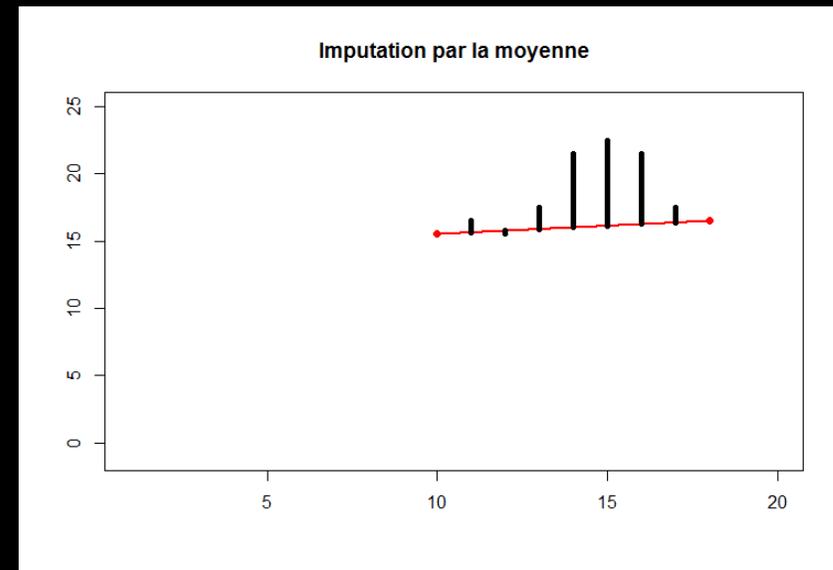
Phase II : ajout d'une variation



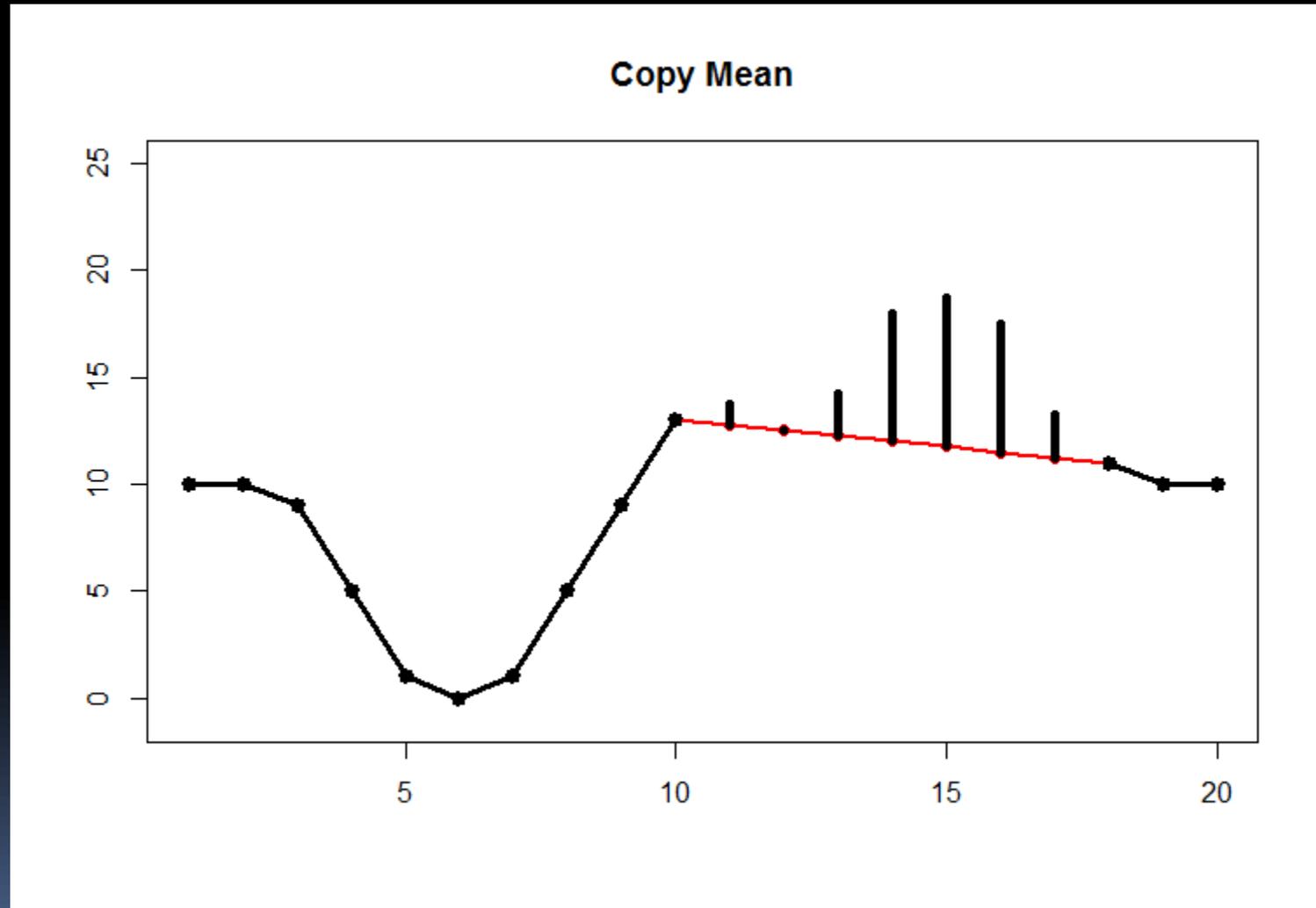
CopyMean = Phase I + Phase II



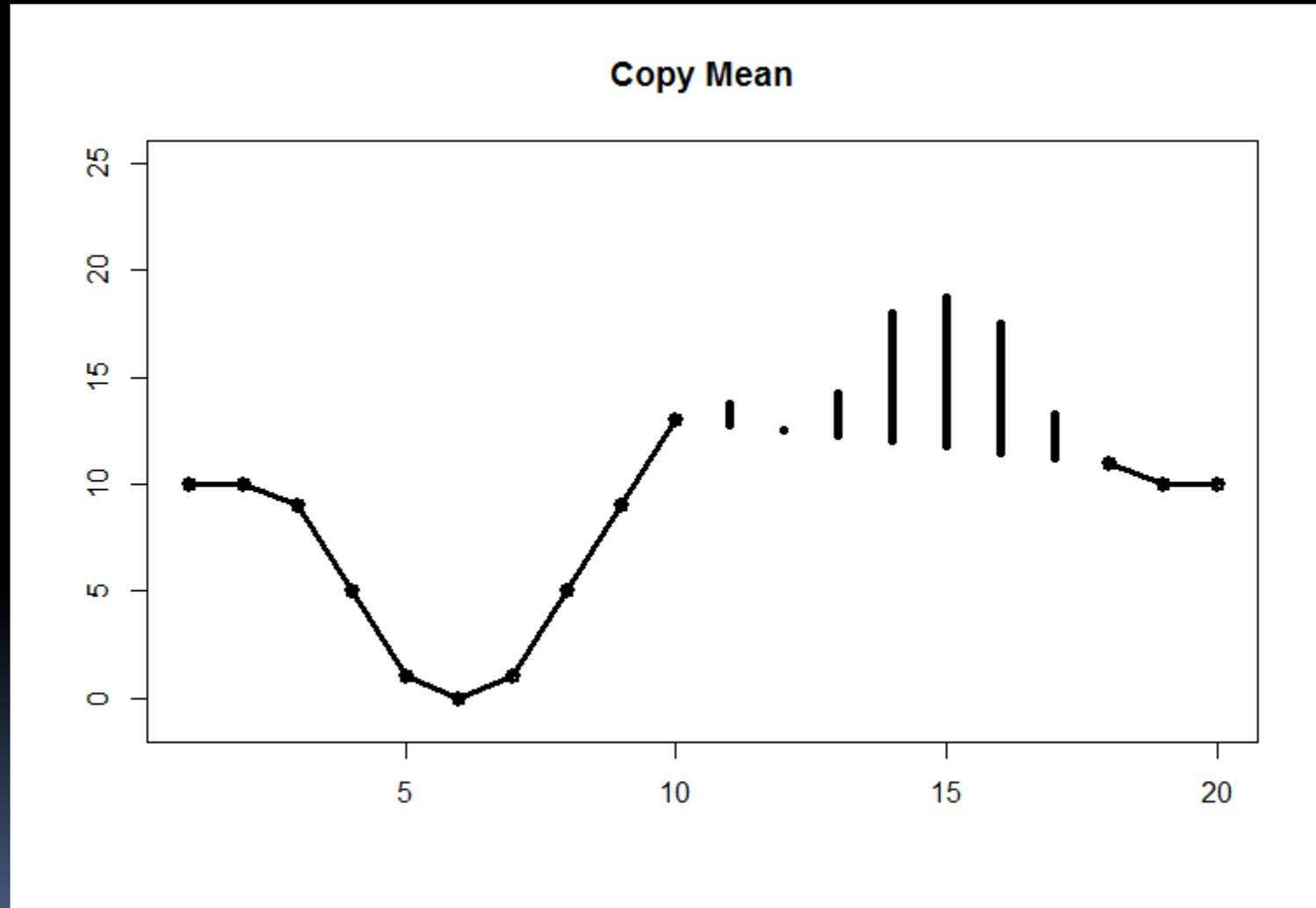
+



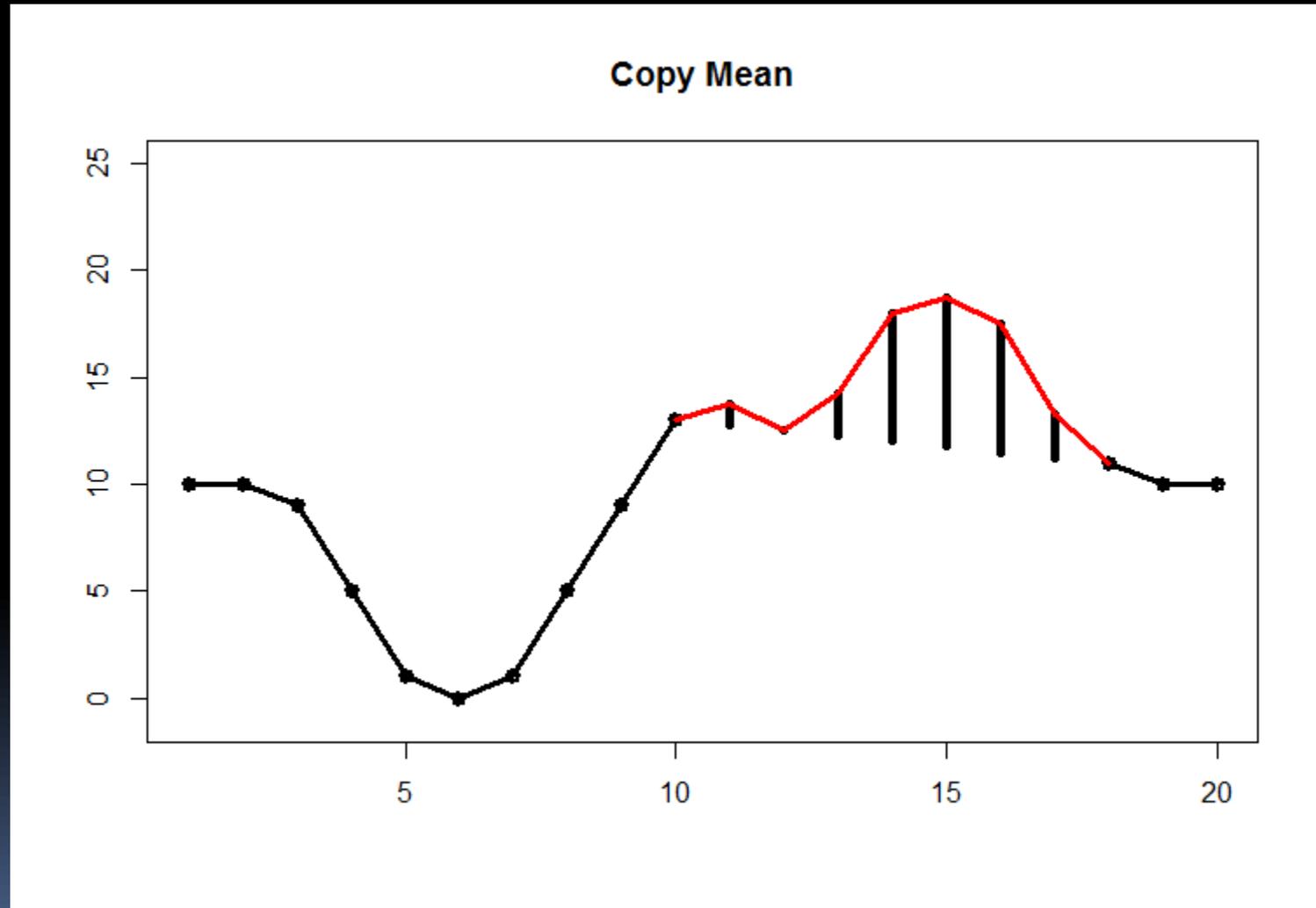
Phase II : ajout d'une variation



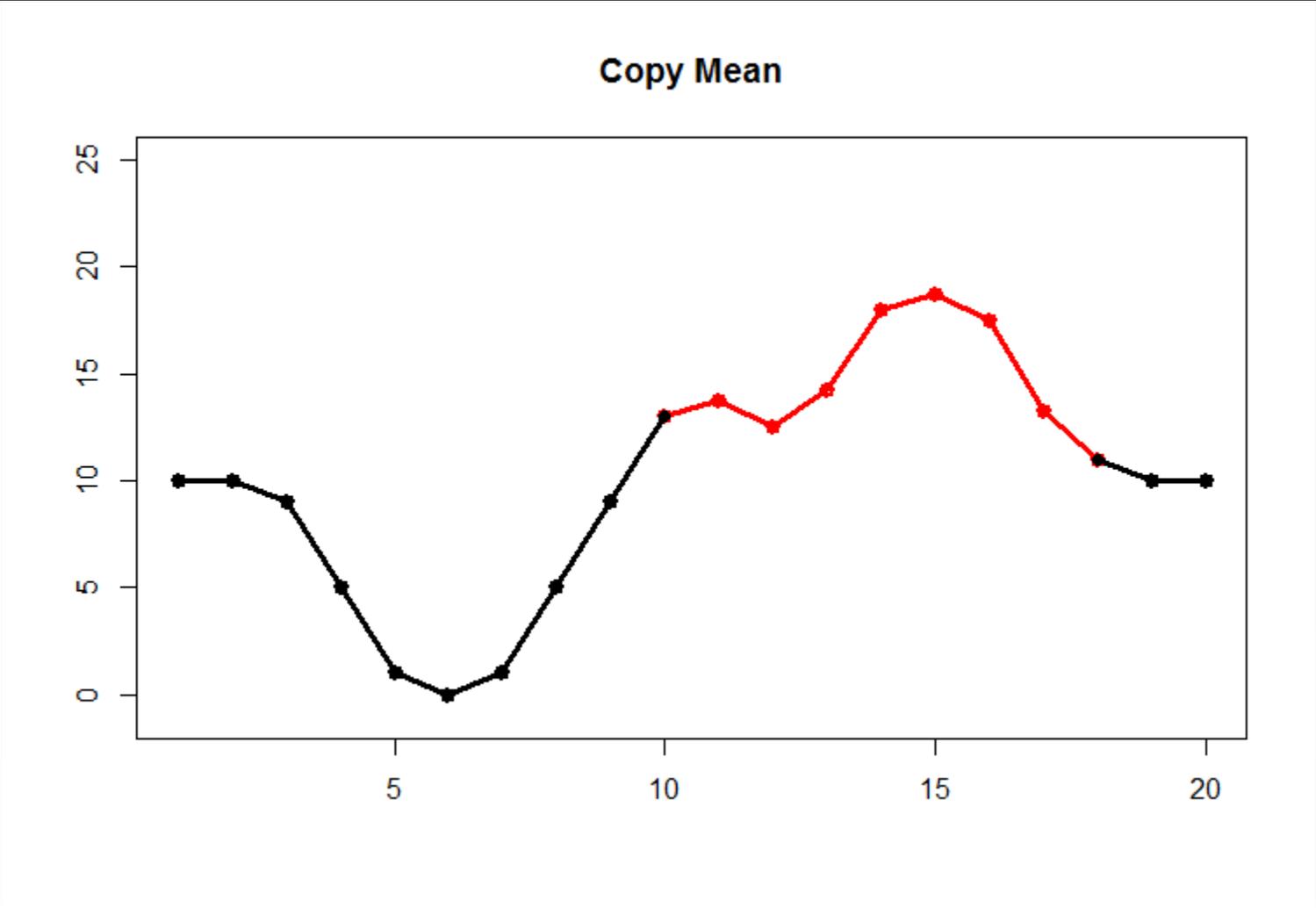
Phase II : ajout d'une variation



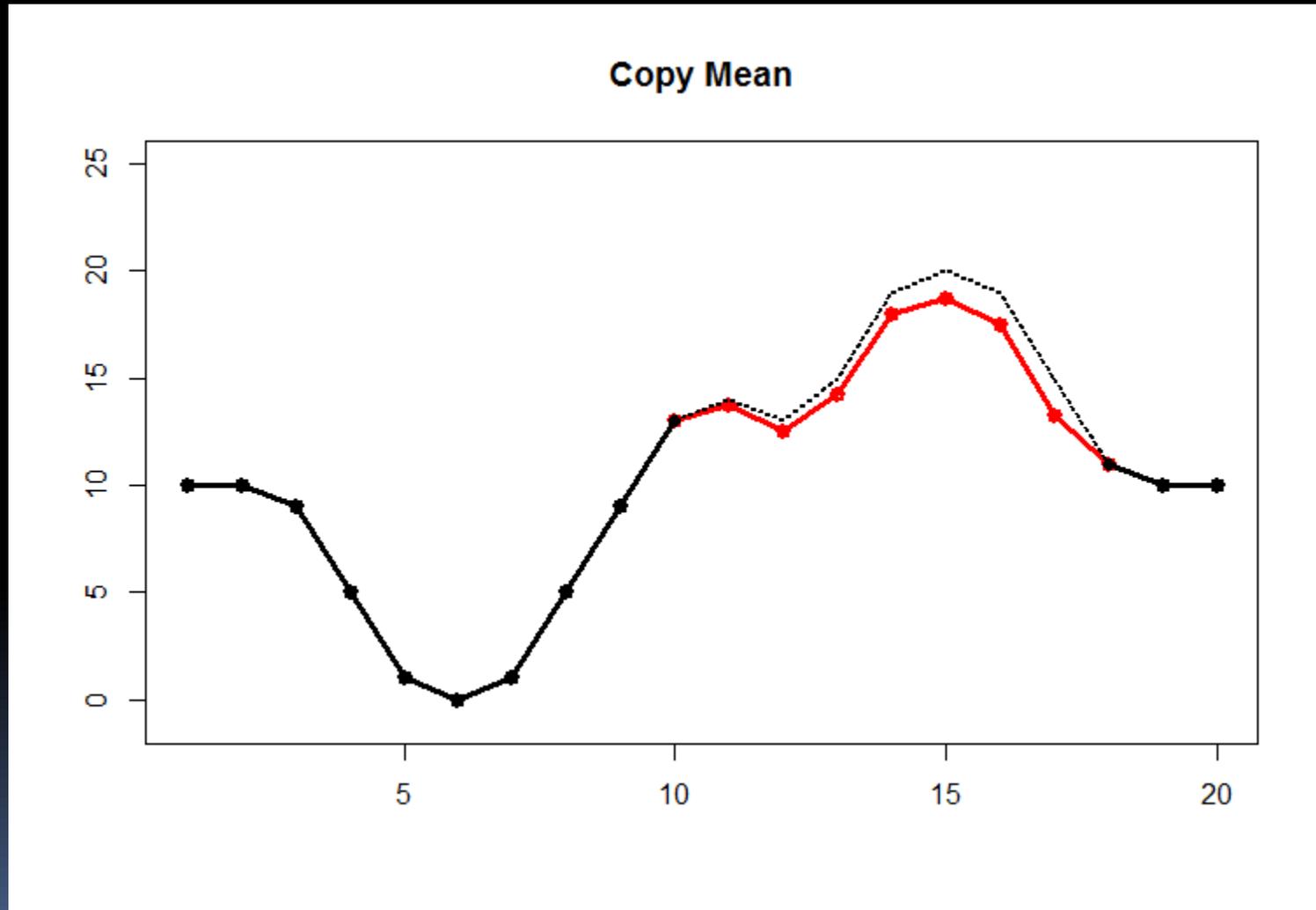
Phase II : ajout d'une variation



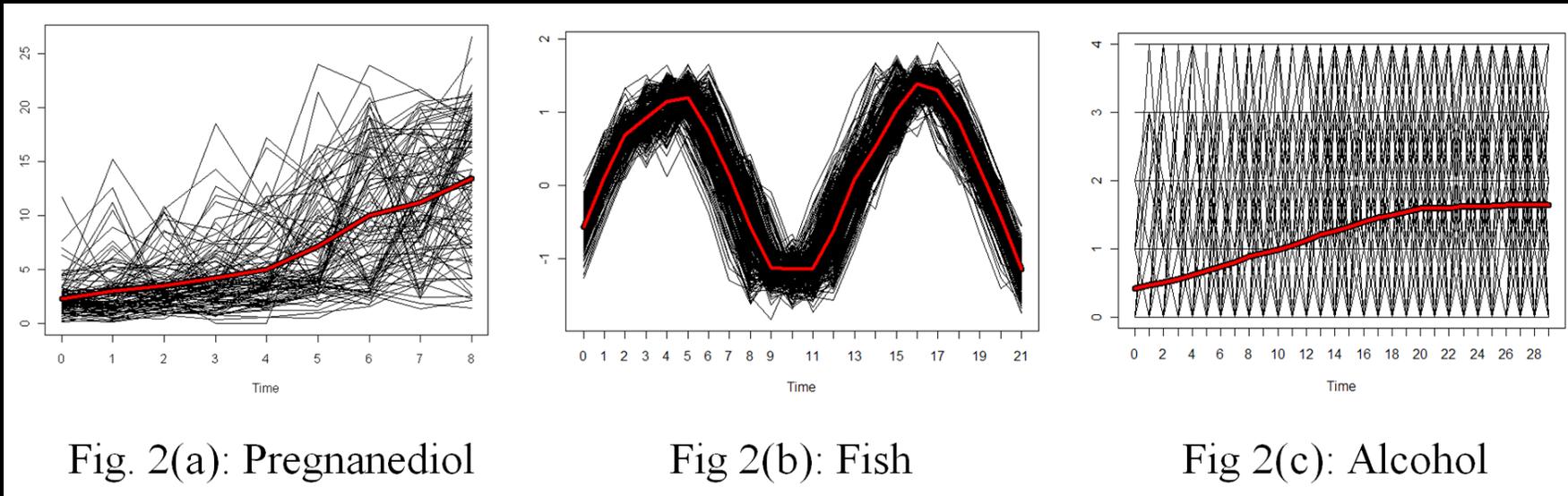
copyMean



copyMean



Exemples réels

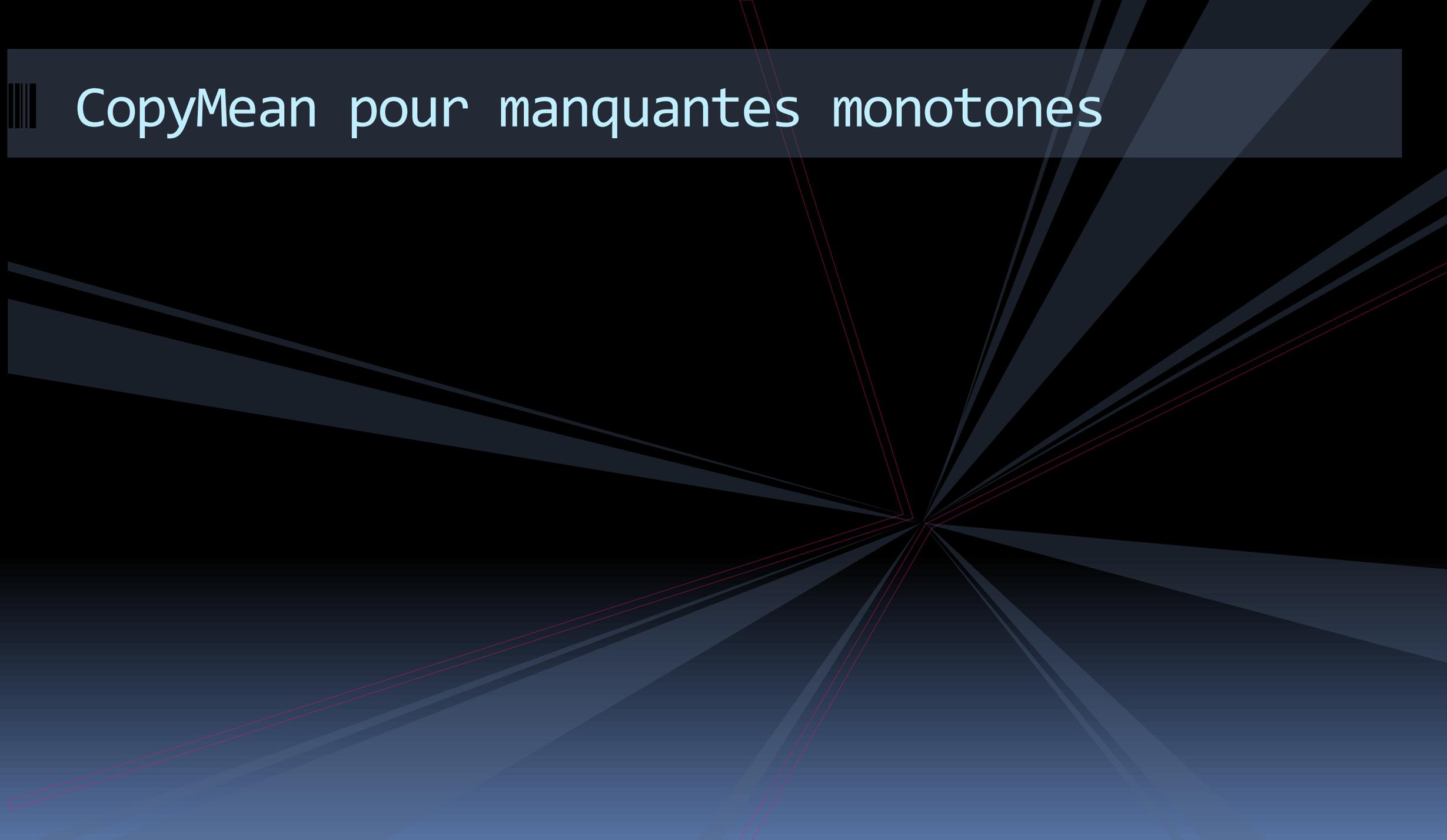


Résultats

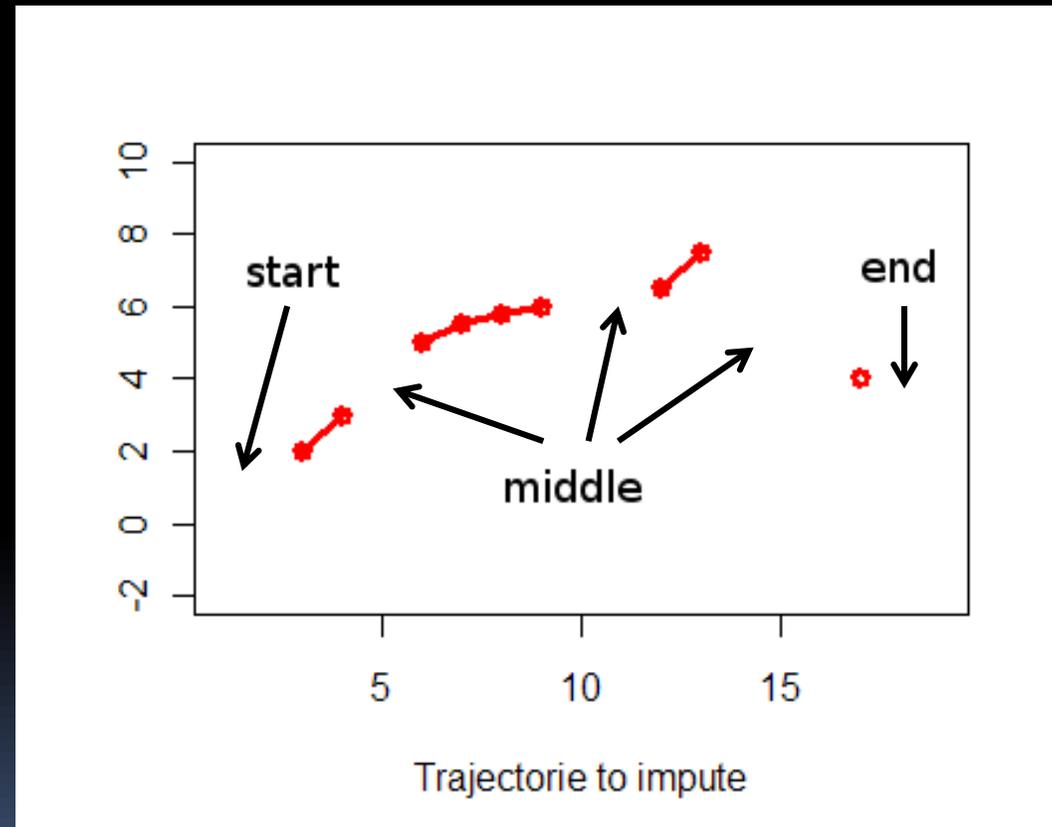
Méthode d'imputation	Nombre d'occurrences ou une méthode est « performante » (sur 27)
1. CrossMean	2
2. CrossMedian	7
3. CrossHotDeck	6
4. TrajMean	6
5. TrajMedian	5
6. TrajHotDeck	3
7. LOCF	10
8. LinearInterpol	17
9. Spline	6
10. CopyMean	26
11. RegressionInt	11
12. RegressionExt	5



CopyMean pour manquantes monotones

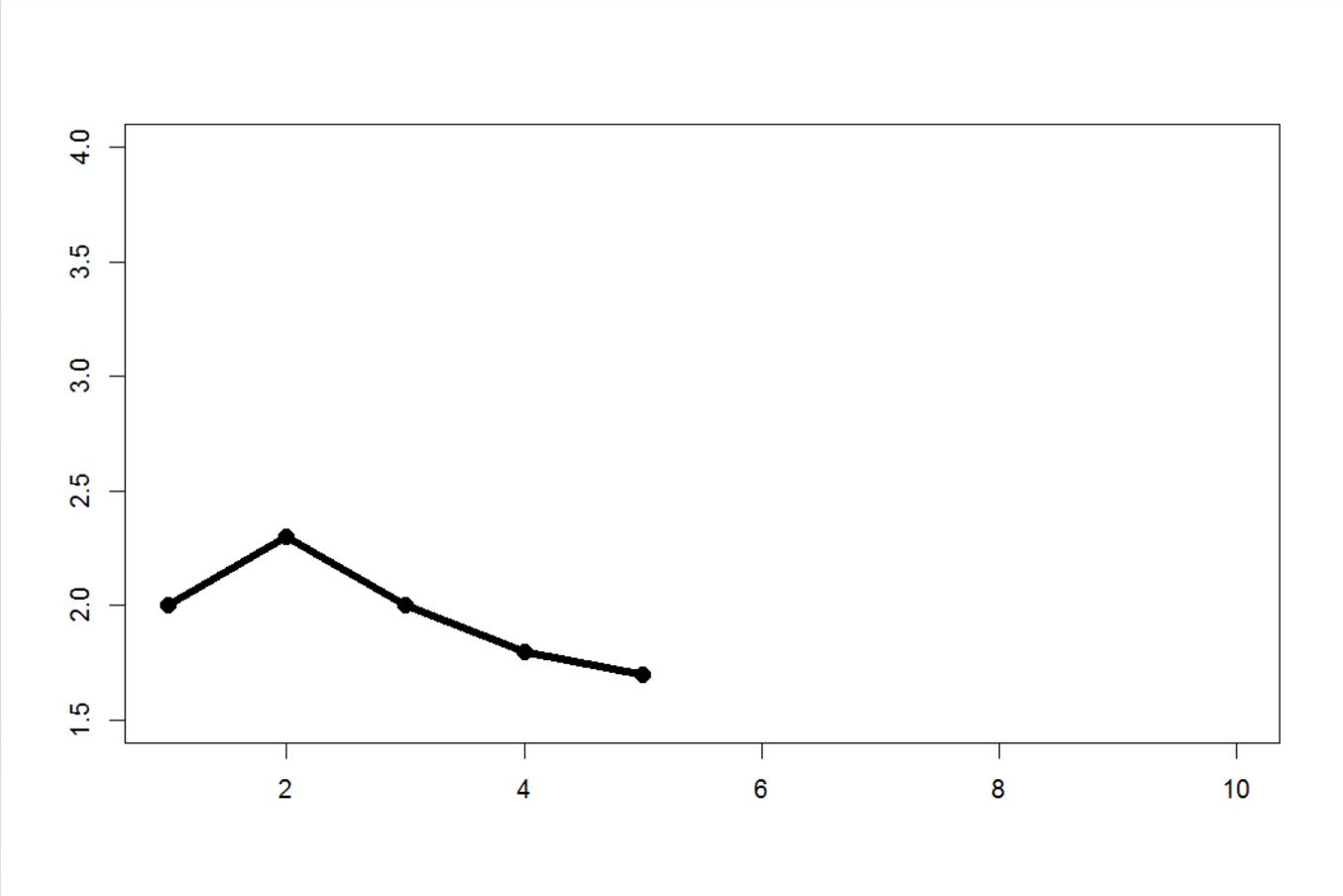


Classification des manquantes



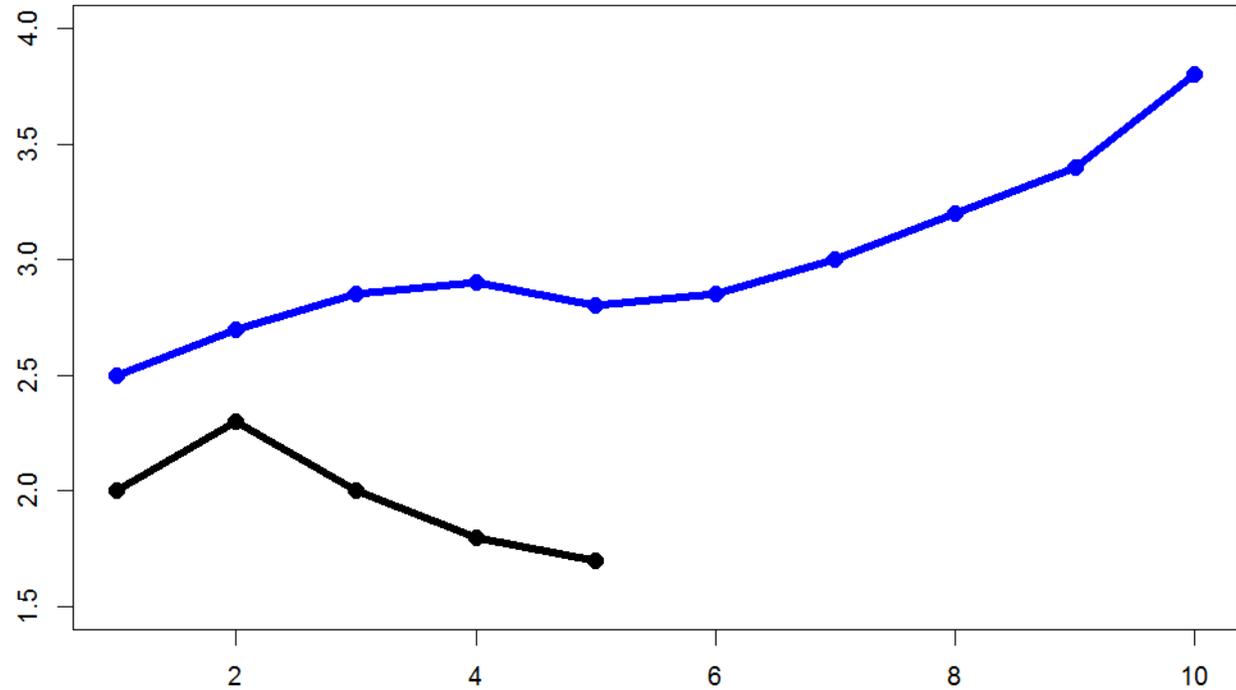


CopyMean pour manquantes monotones



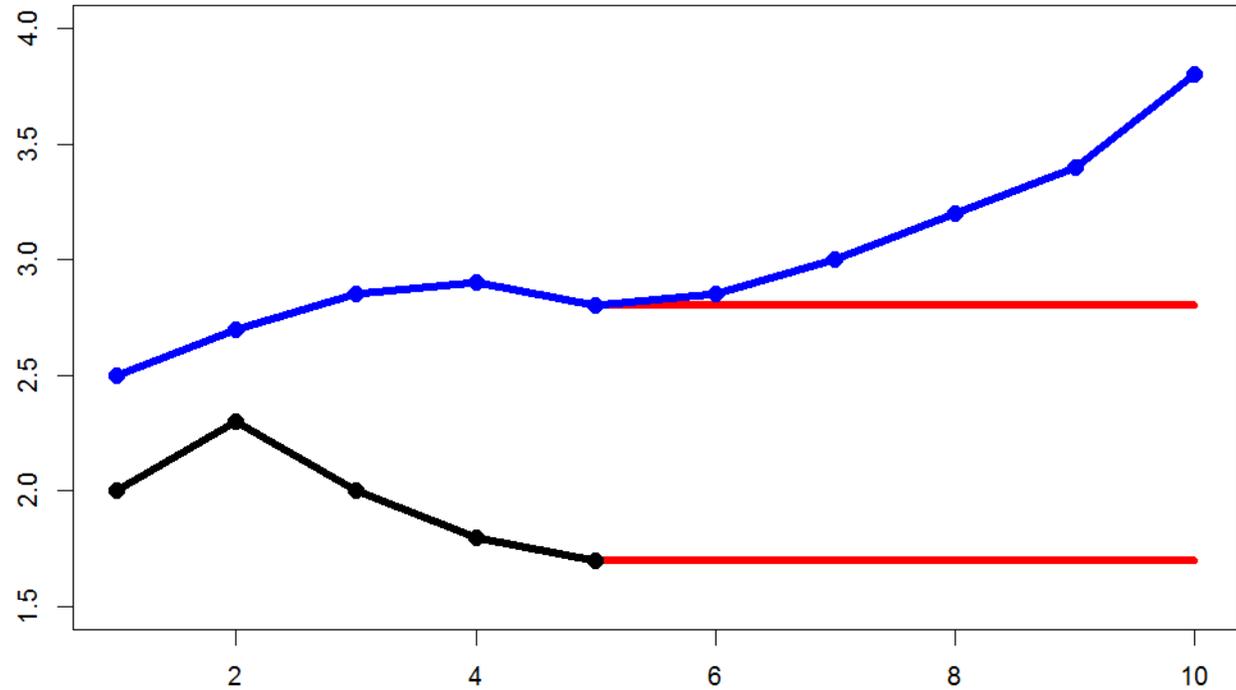


CopyMean pour manquantes monotones



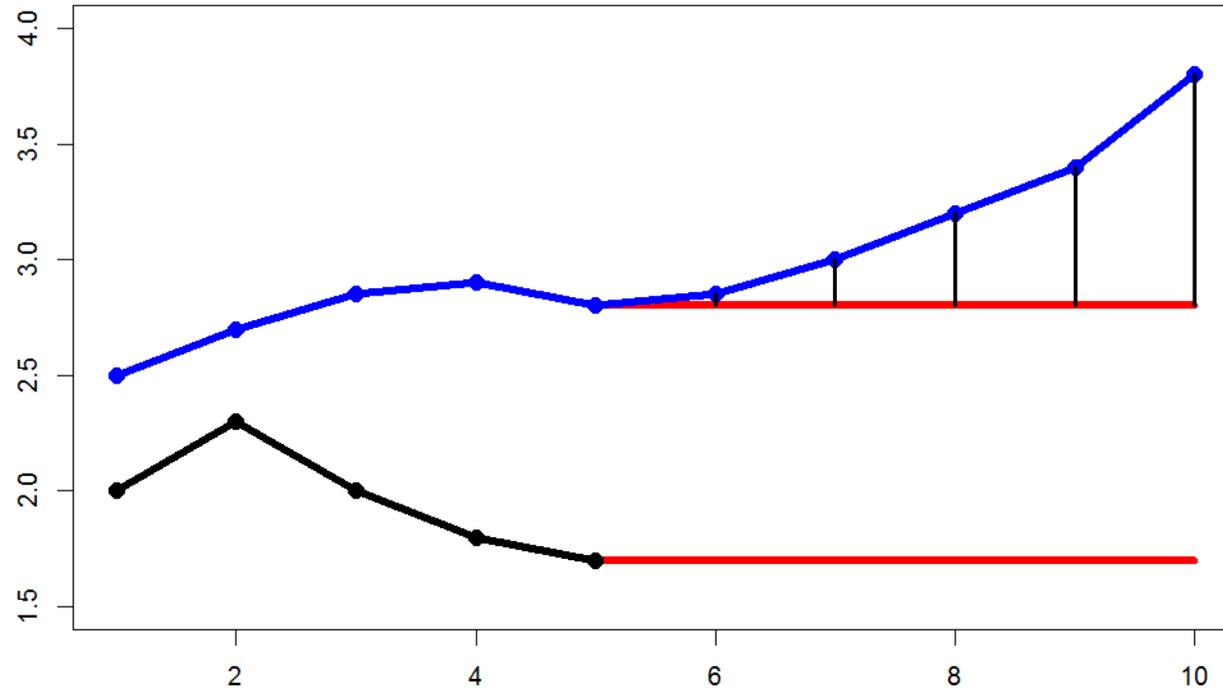


CopyMean pour manquantes monotones



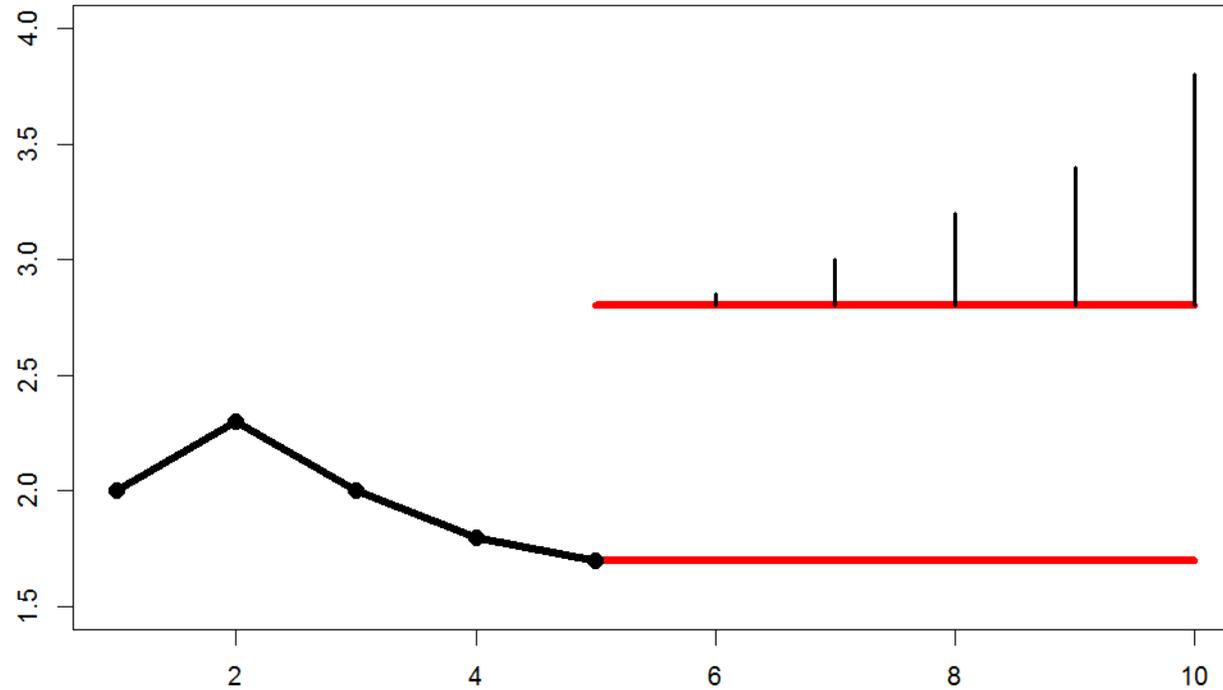


CopyMean pour manquantes monotones



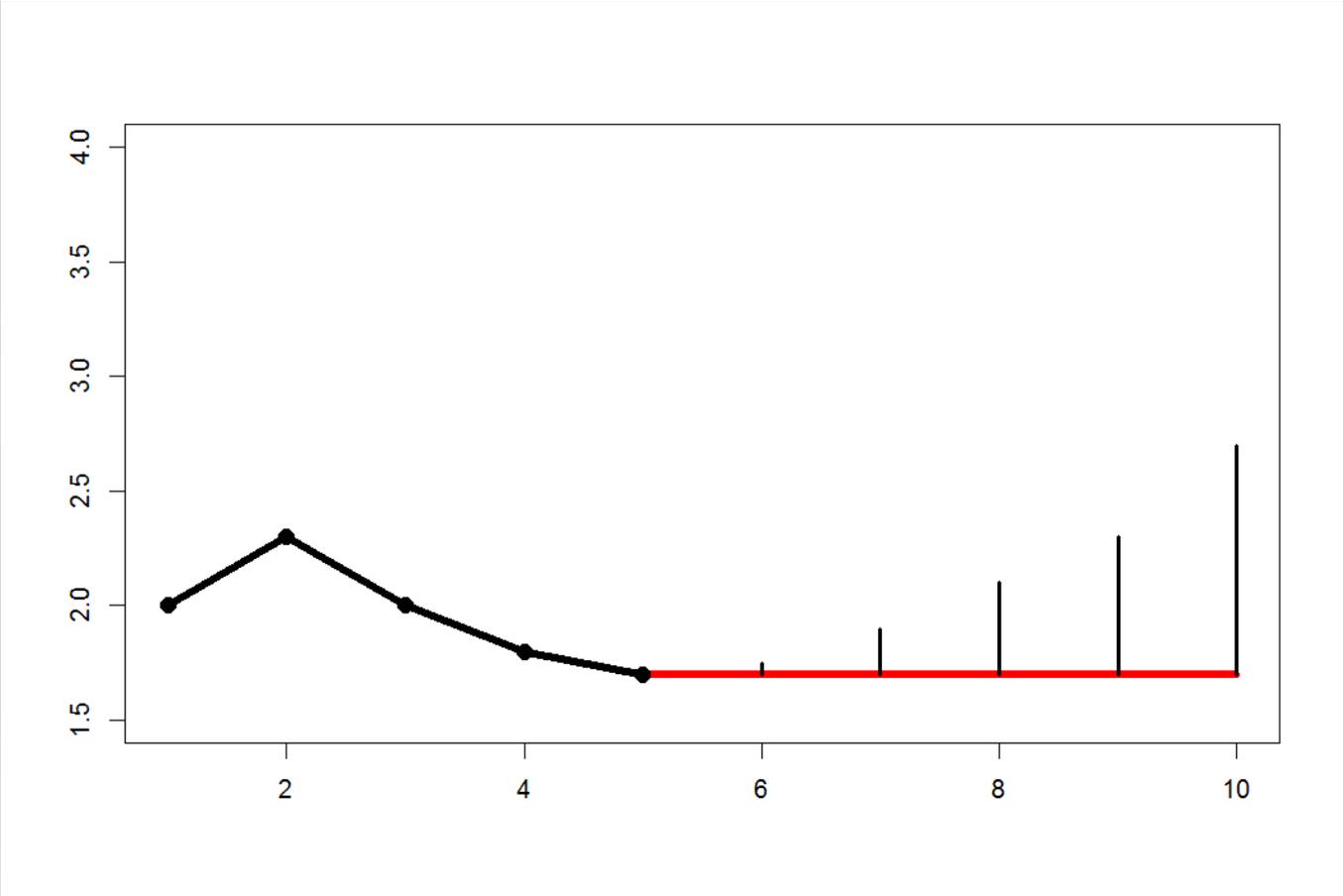


CopyMean pour manquantes monotones



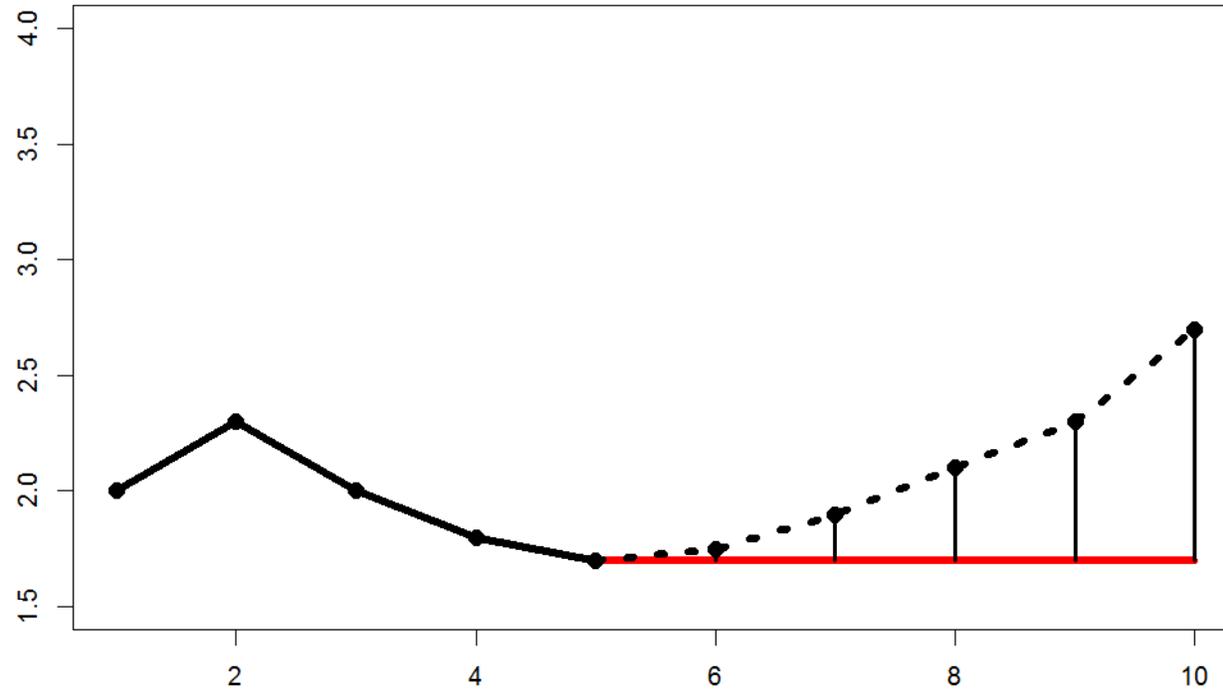


CopyMean pour manquantes monotones



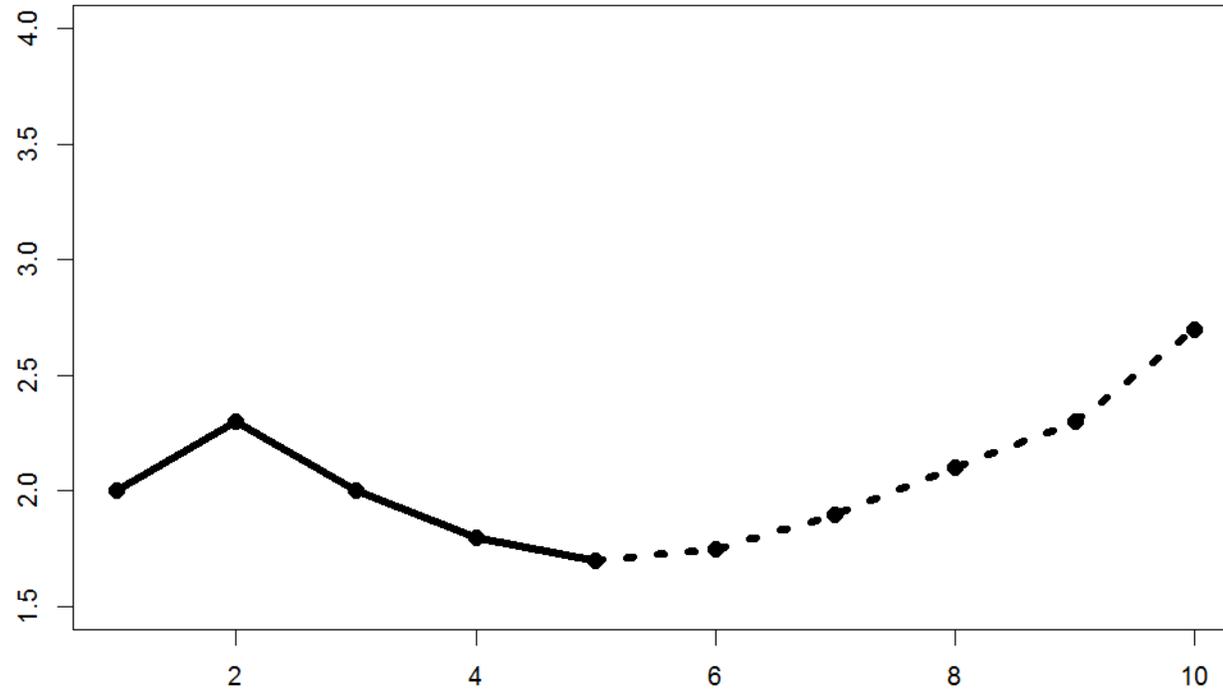


CopyMean pour manquantes monotones





CopyMean pour manquantes monotones



Résultats

Imputation method	MAD	RMSD	Bias	Total
1. Cross Mean	0	0	2	2
2. Cross Median	0	0	0	0
3. Cross Hot Deck	0	0	2	2
4. Traj Mean	0	0	0	0
5. Traj Median	0	0	0	0
6. Traj Hot Deck	0	0	0	0
7. LOCF	7	0	0	7
8. Interpolation <i>Global</i>	0	0	1	1
9. Interpolation <i>Local</i>	0	0	0	0
10. Interpolation <i>Bisector</i>	0	0	2	2
11. Spline Interpolation	0	0	4	4
12. Copy Mean <i>LOCF</i>	11	18	5	34
13. Copy Mean <i>Global</i>	0	0	1	1
14. Copy Mean <i>Local</i>	0	0	0	0
15. Copy Mean <i>Bisector</i>	0	0	1	1
16. Linear Regression,Internal	0	0	0	0

copyMean sous R

- Librairie « kml »



Merci de votre attention

Des questions ?



THE NEXT STEP



www.rplusplus.org