

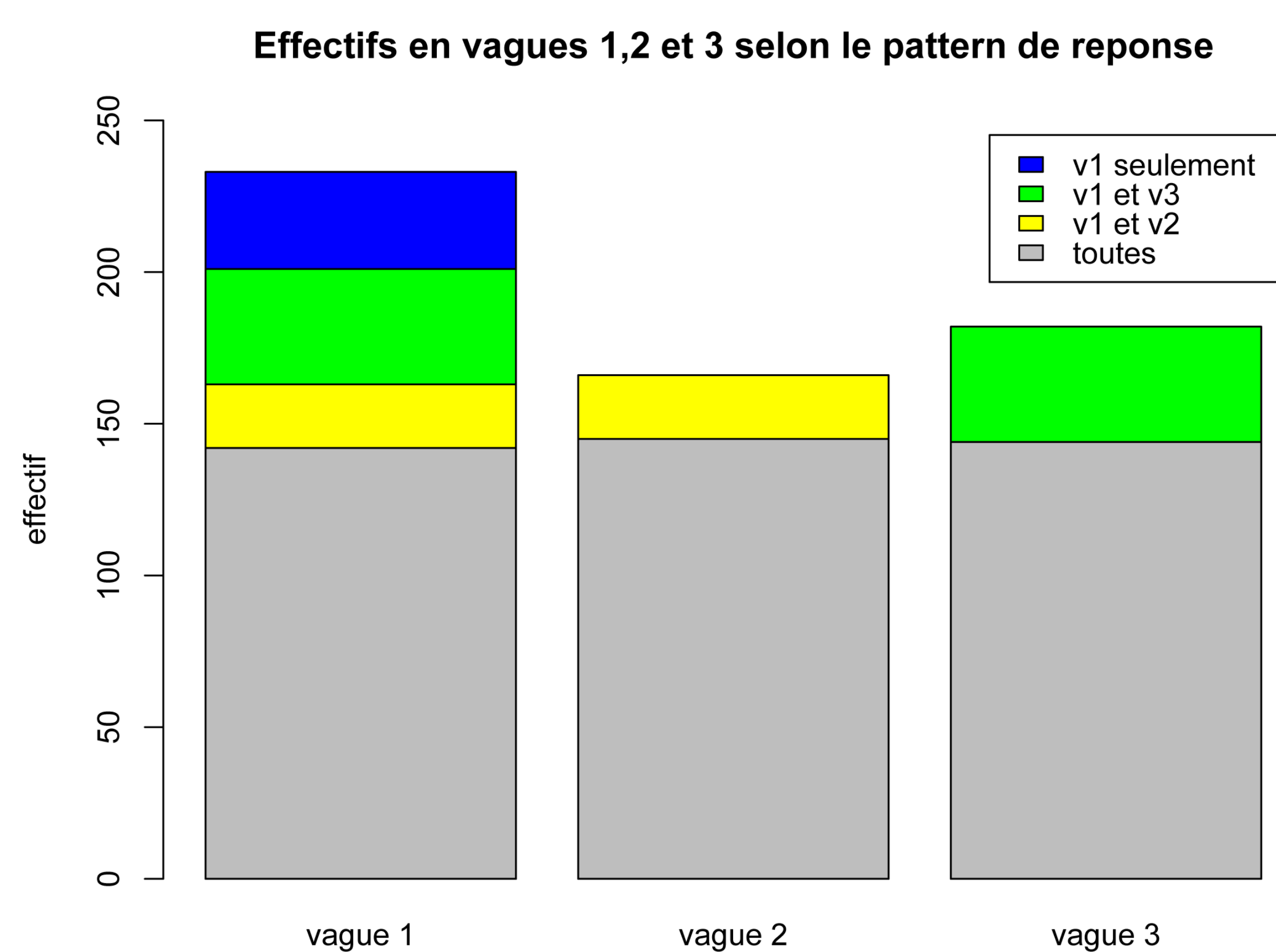
Introduction et motivation

Les données d'enquête sont fréquemment parsemées de données manquantes (NA) :

	sexe	A	B	C	D	E	F
1	h	NA	3	NA	NA	8	5
2	h	7	NA	NA	18	3	NA
3	h	5	1	NA	25	1	8
4	f	8	2	NA	82	4	7
5	f	9	1	72	63	3	6

Nous étudions les méthodes de traitement des données manquantes **NA** et cherchons des solutions performantes, notamment dans le cas longitudinal.

Notre travail est motivé par l'enquête "devenir parent" menée par le centre lémanique d'étude des parcours et modes de vie (PAVIE). Sur 236 couples ayant participé en vague 1, seulement 166 ont répondu à l'appel en vague 2 et 182 lors des derniers entretiens.



Procédure

Nous avons comparé plusieurs méthodes de traitement de données manquantes sur des données catégorielles longitudinales et avons estimé un modèle logistique. Nous nous sommes intéressés à la distribution des variables et à l'estimation du modèle.

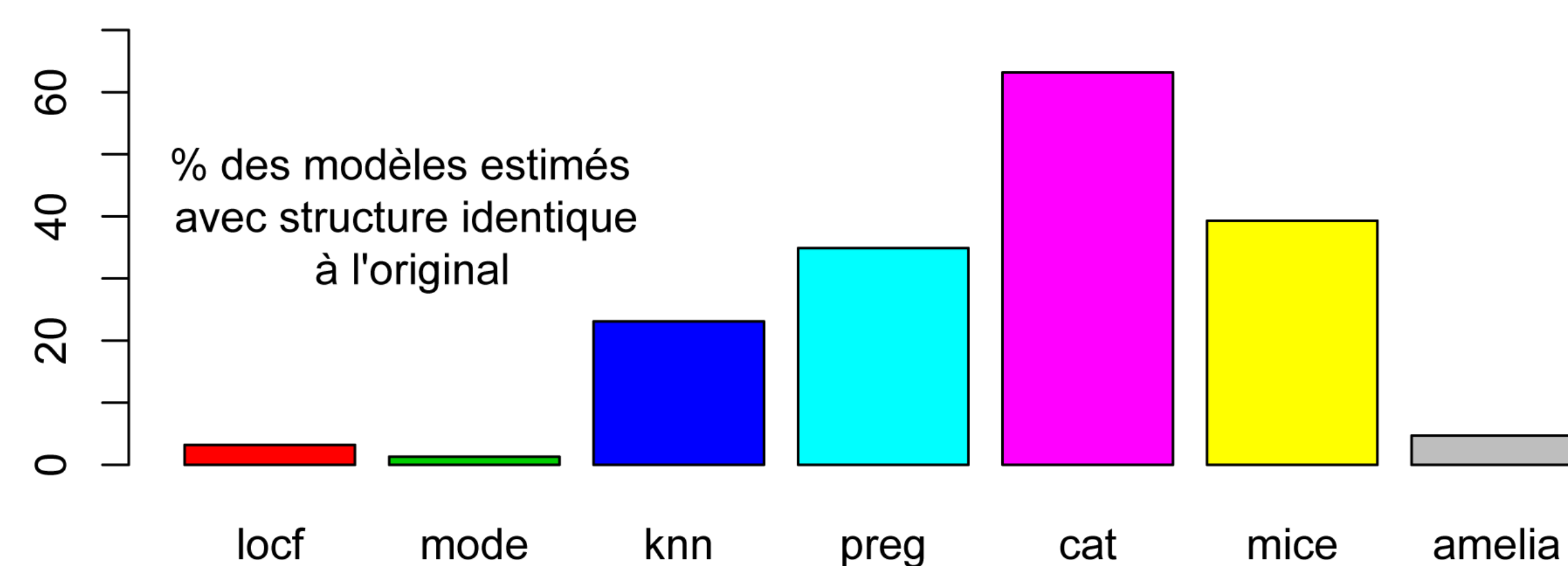
Résultats

Les méthodes préservant le mieux les distributions univariées et bivariées sont **preg** (imputation simple basée sur la régression logistique), **mice** (imputation multiple basée sur la régression logistique) et **cat** (imputation simple pour variables catégorielles basée sur l'algorithme EM et modèle loglinéaire).

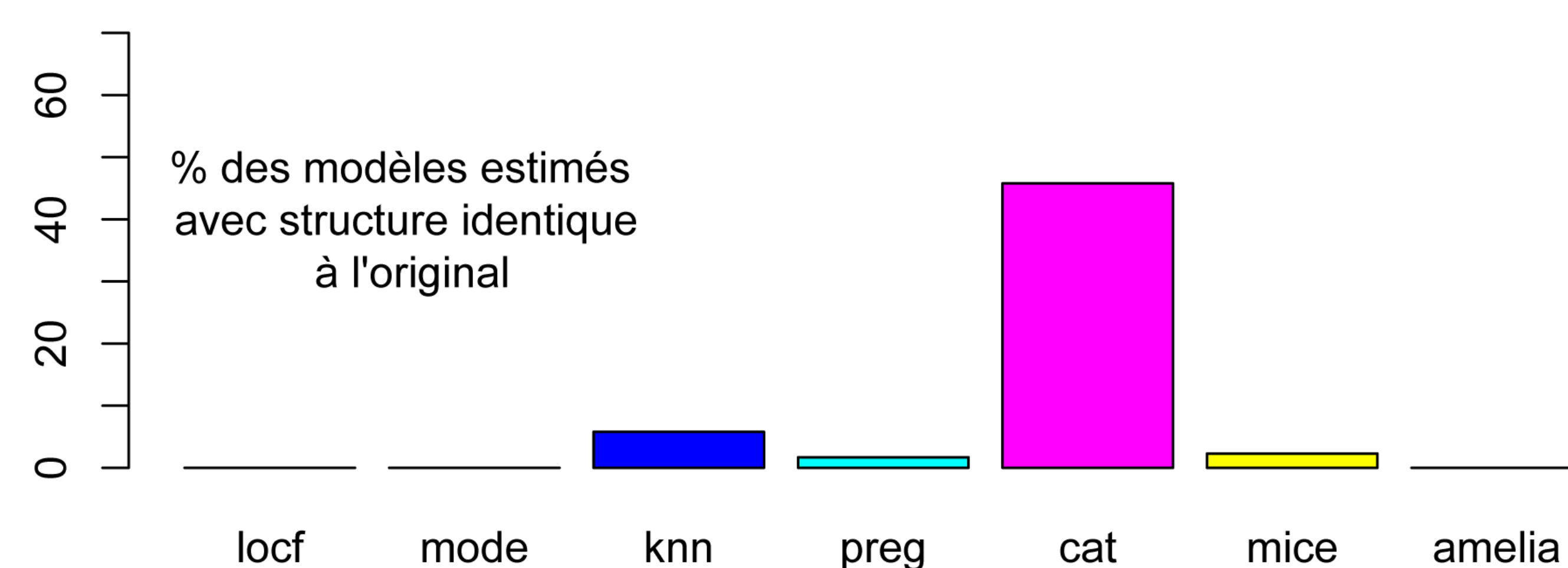
Pour l'estimation du modèle, cette tendance reste perceptible mais la méthode **cat** se distingue des autres.

Lorsque les NA sont distribuées sur trois variables seulement (graphique du haut) et en particulier sur la variable expliquée (bas), la méthode **cat** préserve le mieux la structure du modèle. C'est la seule méthode qui permet de retrouver les mêmes variables significatives que dans le modèle original dans au moins 45 % des cas quelque soit la situation.

3 variables avec NA: deux temporelles et une covariable.



3 variables avec NA: deux temporelles et la variable expliquée.



Les estimations des coefficients de régression logistique par la méthode **cat** ne sont pas toujours les plus précises (dispersion parfois plus élevée que d'autres) mais sont les moins biaisées.