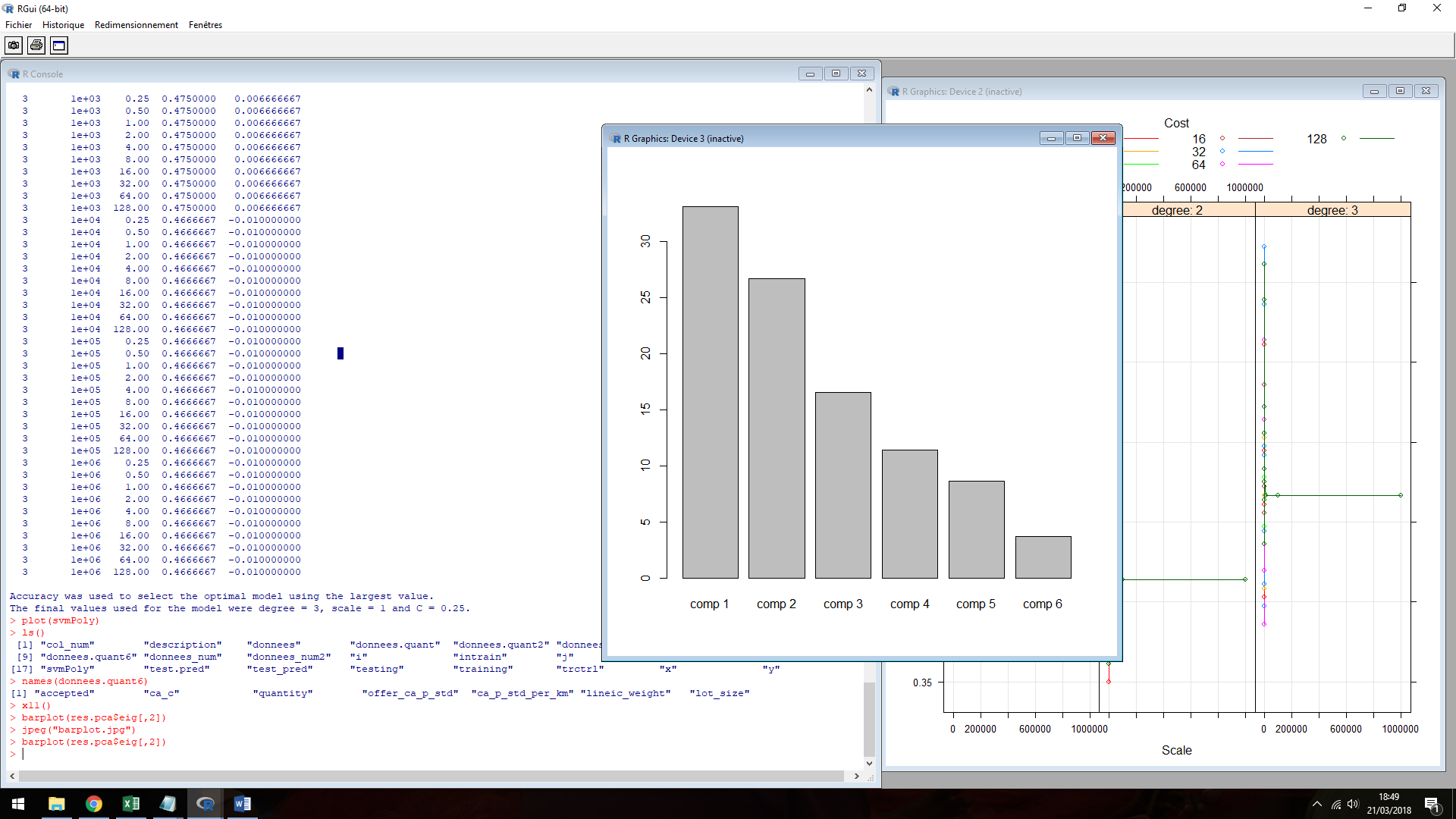
## ACP

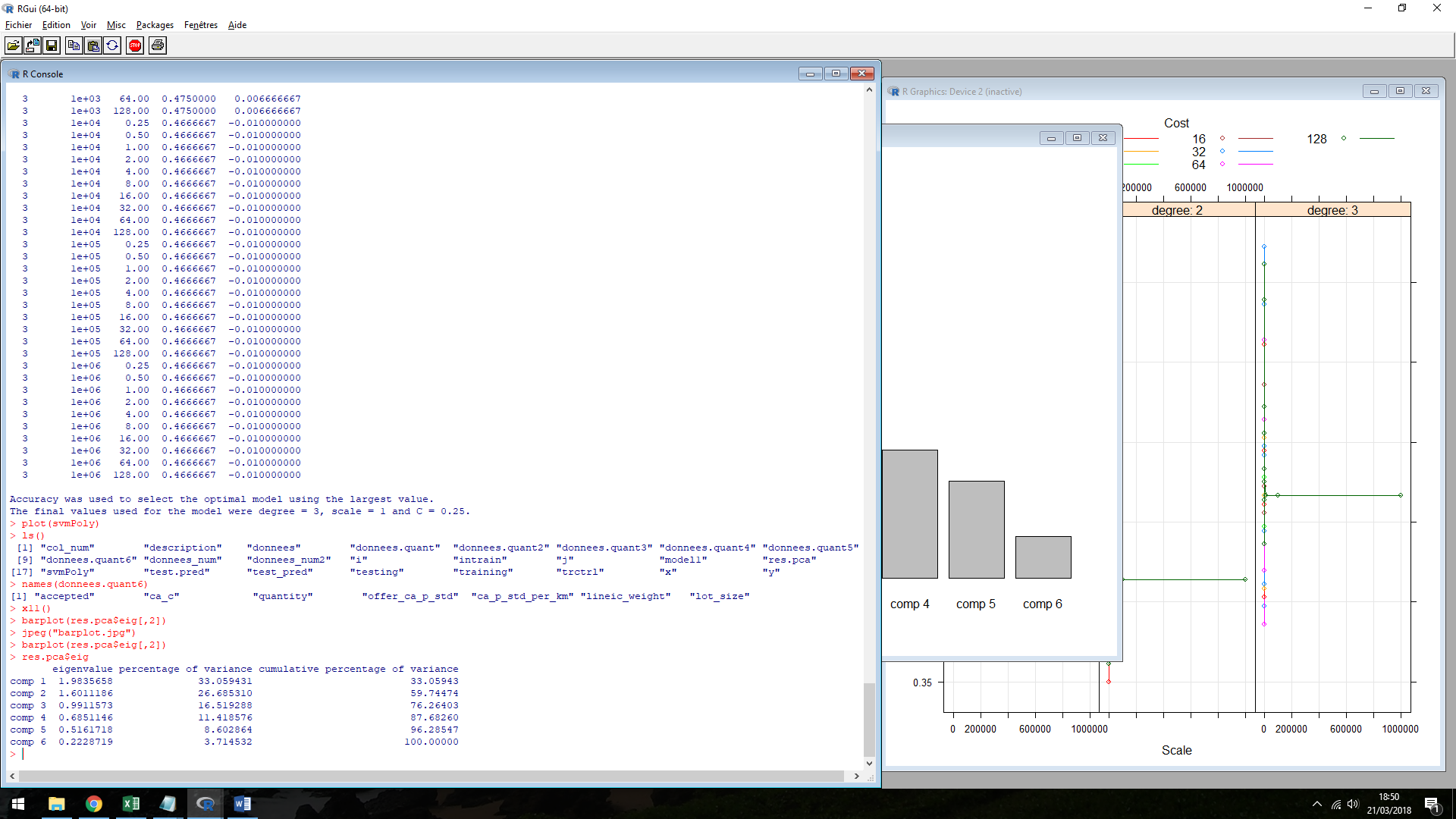
J’ai pris un sous-ensemble des variables quantitatives pour réaliser l’ACP :

ca\_c, quantity, offer\_ca\_p\_std, ca\_p\_std\_per\_km, lineic\_weight et lot\_size.

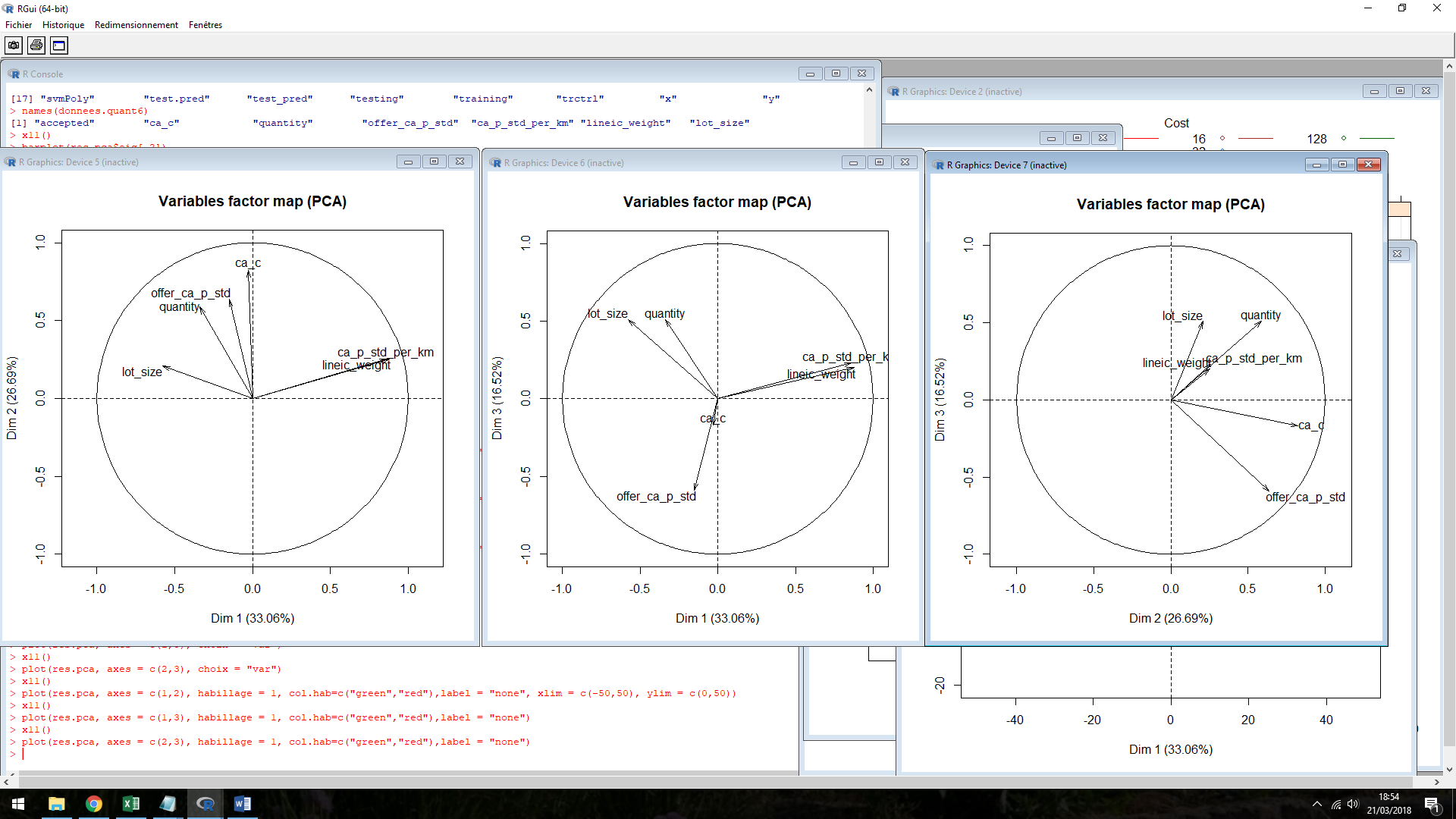
Mon ordi a eu du mal à tenir la longueur, j’ai pris seulement celles dont je comprenais l’intitulé (et qui étaient quantitatives).

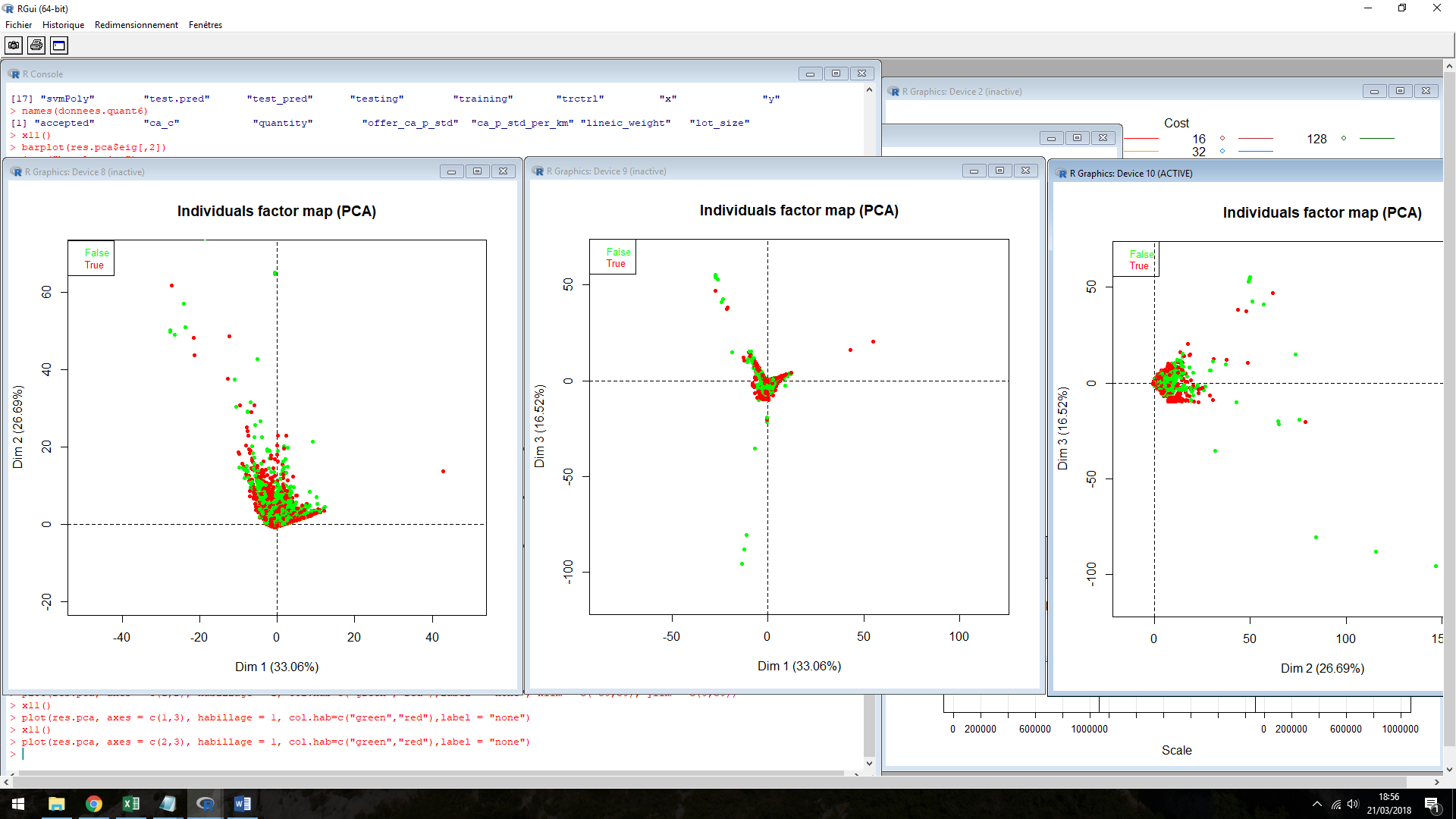
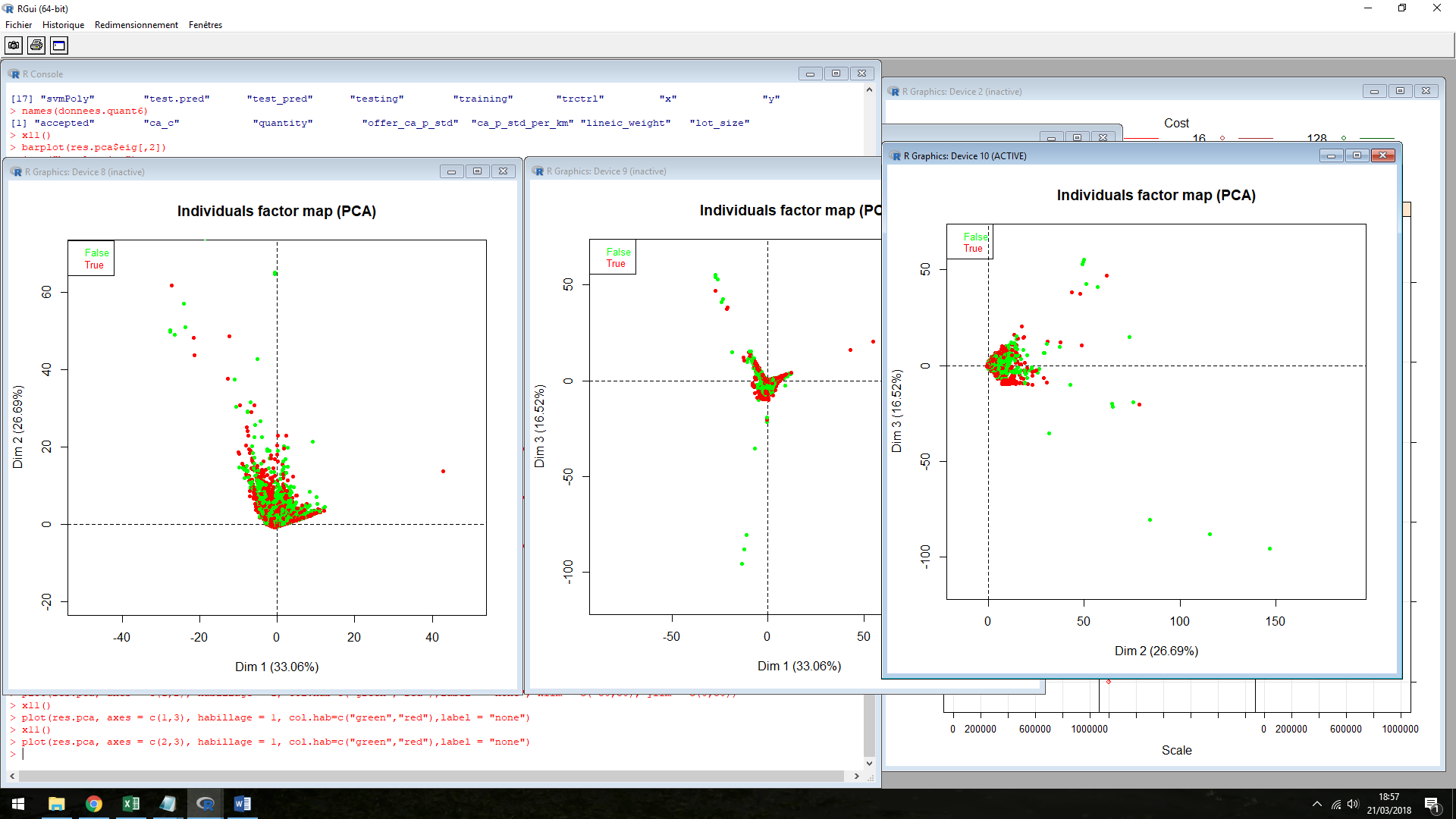
En regardant le barplot on se rend compte qu’il n’y a rien qui puisse nous aider à discriminer le nombre de dimension à sélectionner.





En revanche, d’après la règle de Kaiser on devrait en prendre 2 (strictement). Cependant, la faible quantité d’inertie que l’on trouve dans ces deux dimensions me pousse à en garder une troisième pour avoir + de 70% d’inertie.





On peut voir ici que le cout au km et le poids linéaire sont très corrélés -> risque de suraprentissage ? et ne sont en revanche pas corréles au prix, ni à l’offre.

La dimension 1 va séparer les individus en fonction de ces deux variables majoritairement.

La dimension 2 va séparer les grosses offres des petites (ca\_c, quantity, offer\_ca\_p\_std)

La dimension 3 va séparer les grosses offres qui ont suivie le prix du cuivre standard (offer\_ca\_p\_std)

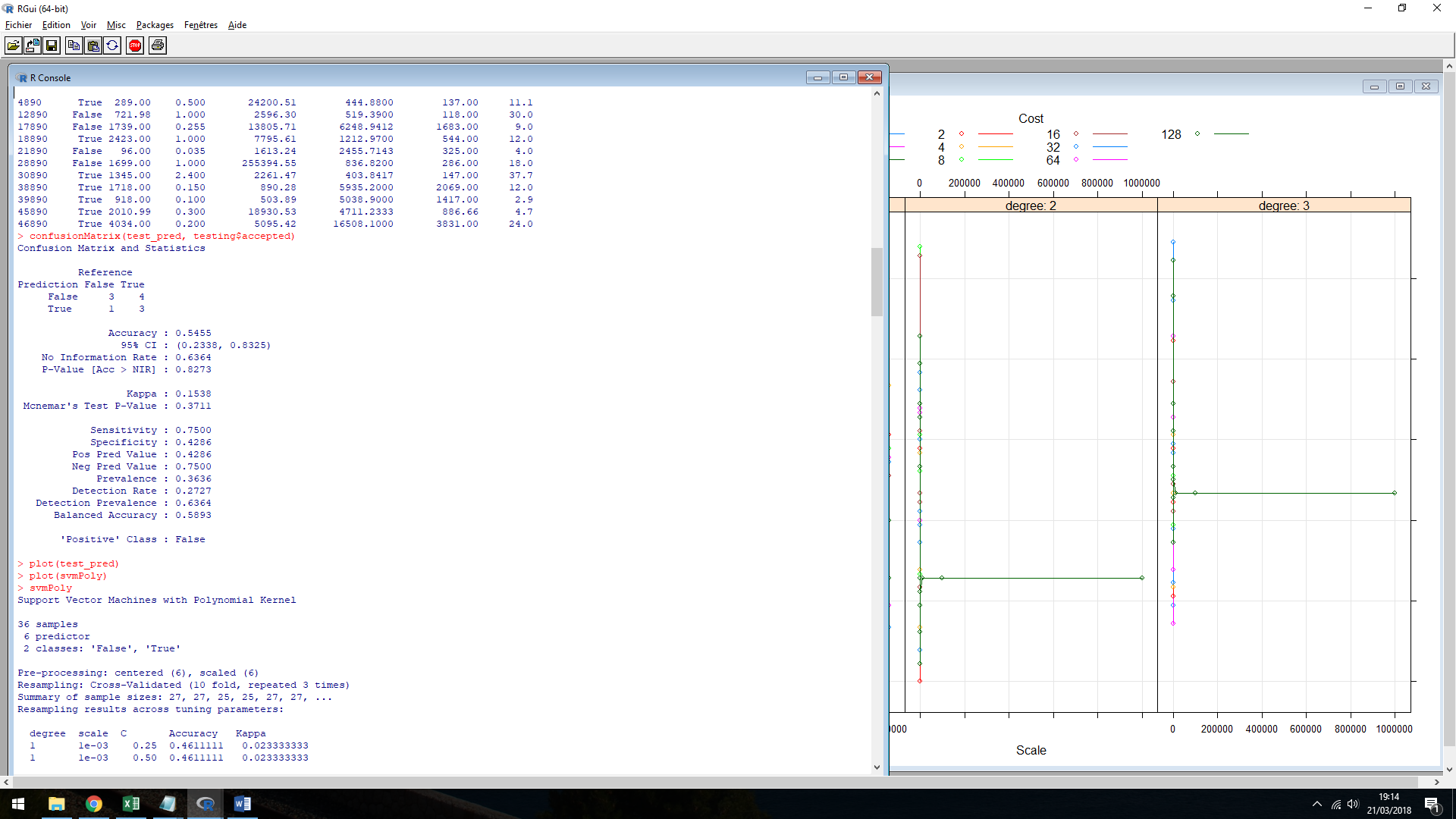
On remarque aussi que les gros contrats sont généralements remportés (dimension 2, ca\_c), et d’autant plus quand on le rapporte au prix du cuivre standard (dimension 3, offer\_ca\_p\_std)

# SVM

J’ai eu beaucoup de soucis avec mon ordi sur les temps de calcul, j’ai dû ramener le data.frame à seulement une cinquantaine de ligne pour faire mon modèle, du surement à une trop grande recherche sur les coefficients.

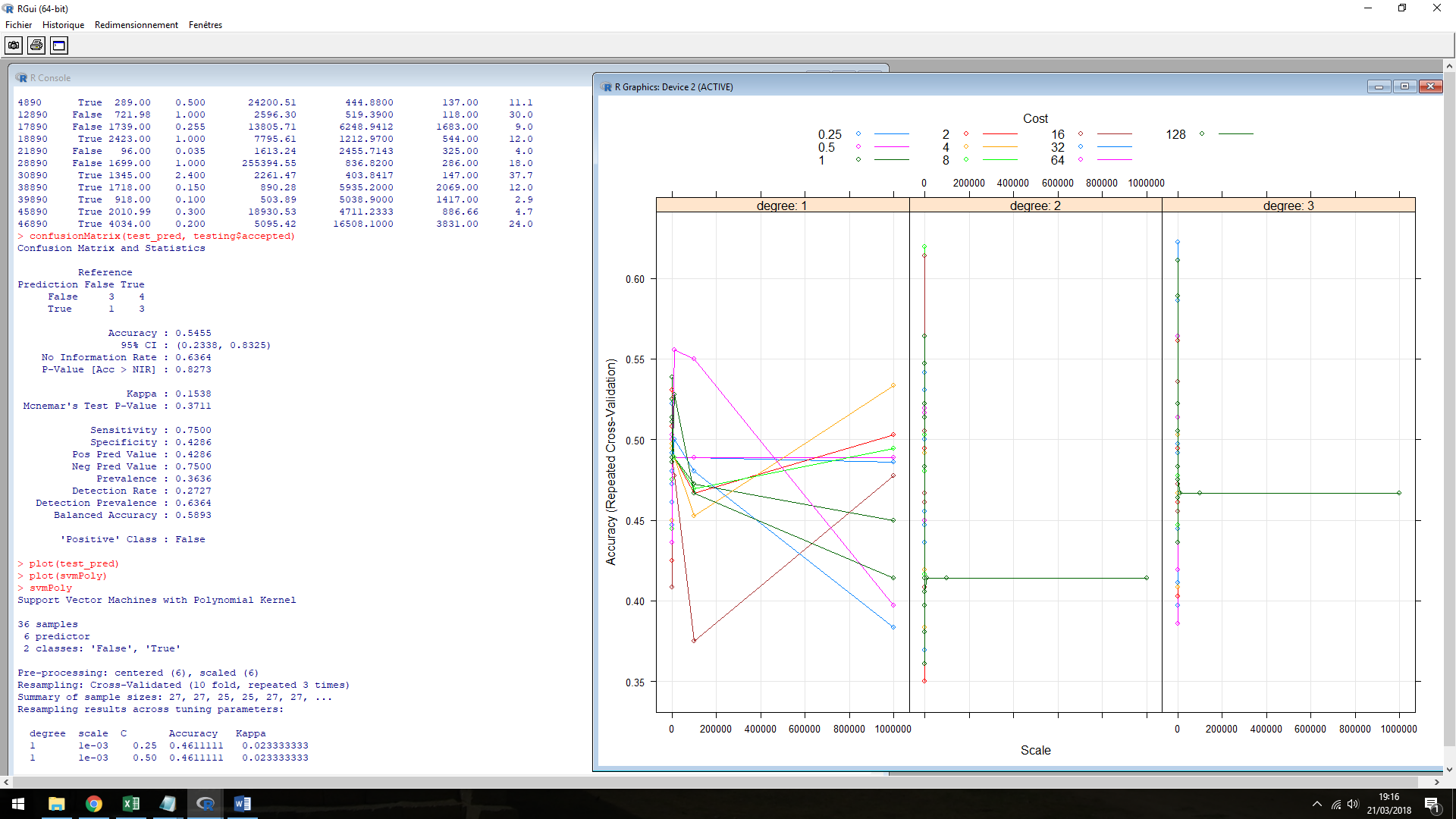
J’ai fait du 10-fold cross-validation pour l’apprentissage, répété trois fois.

La matrice de confusion :



On trouve une sensibilité qui n’est pas pire mais qui vaut toujours mieux que la spécificité.

Pour le choix du modèle, je n’aurais pas choisi celui que « caret » a choisi par defaut car parmi tous ceux qu’il a fait on en trouve avec une accuracy supérieure à 0,6, or sur ce modèle elle est de 0,5455.



Les différents runs pour l’optimisation des paramètres.

Le choix s’est porté sur le param degree = 3, scale = 1 et C = 0.25

Je relancerai bien cette optimisation des paramètres avec un échantillonage pour scale autour de l’unité (entre 0 et 10) plutôt que de monter dans le million.

# Impact du cours du métal

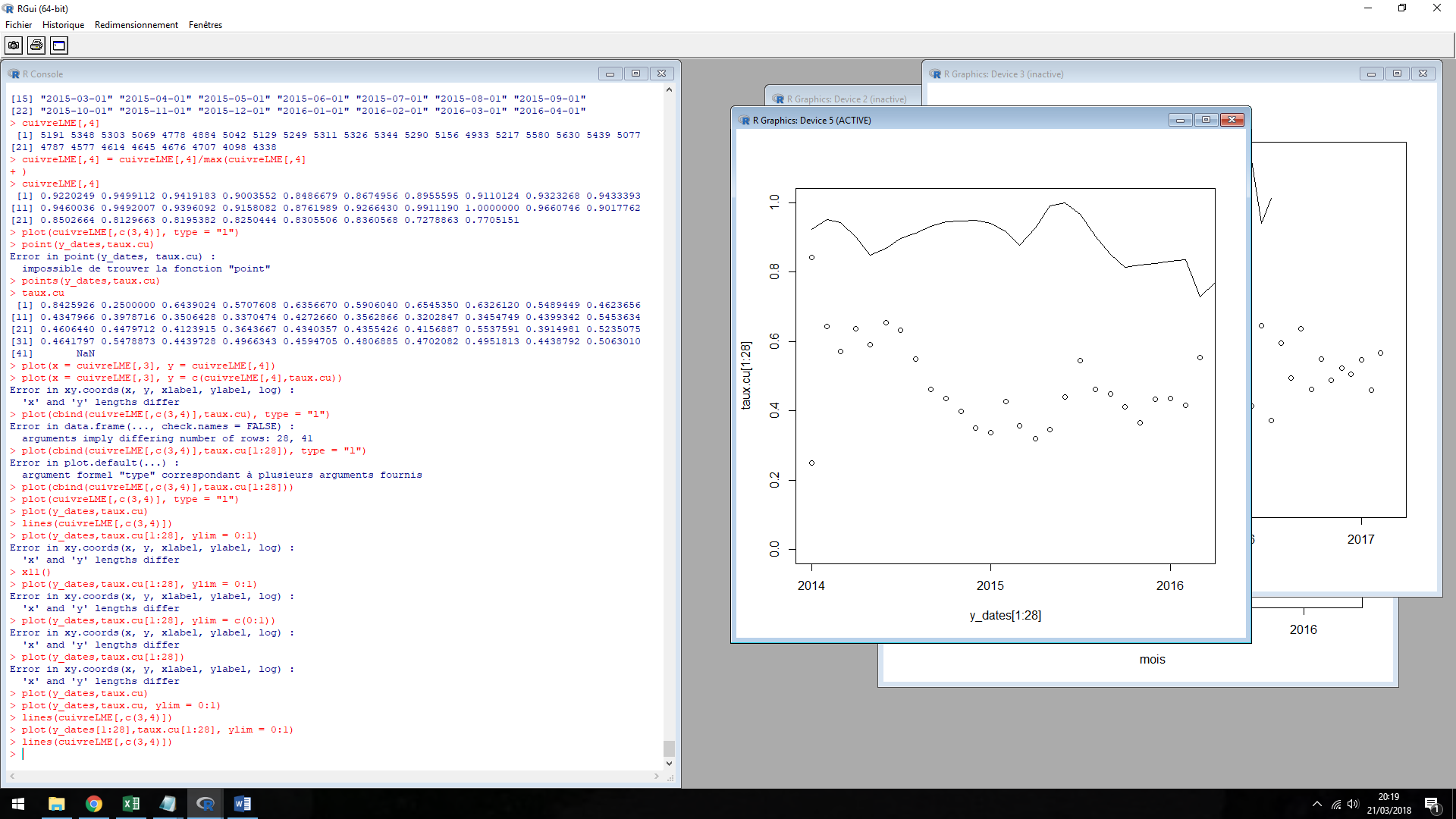
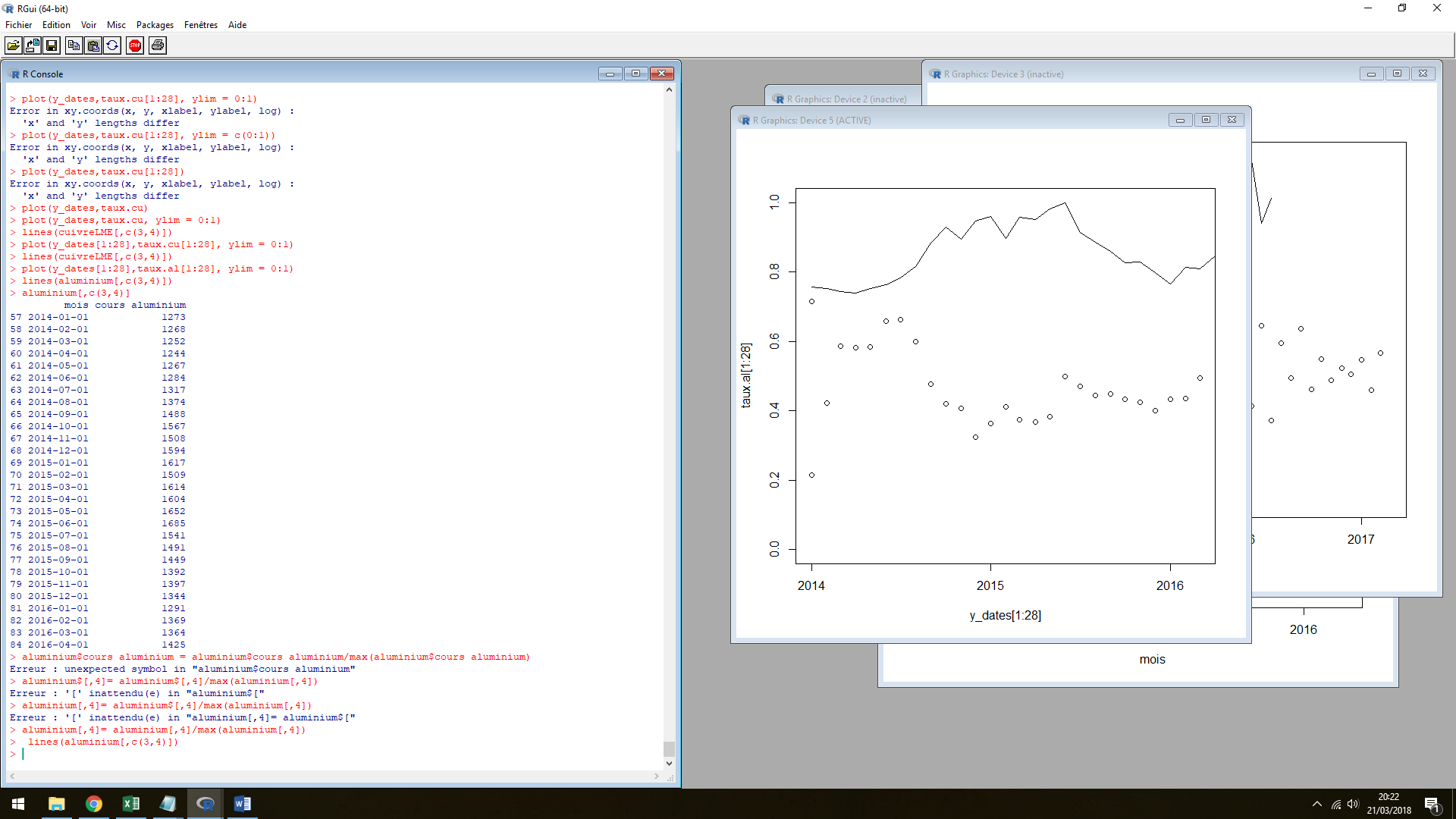
J’ai voulu regarder la corrélation entre le taux de devis accepté et le cours du cuivre/aluminium.

Je trouve des corrélations qui dans les deux cas ne sont pas satisfaisantes :

-0.1233047 pour le cuivre et -0.4031402 pour l’aluminium

Je me demande si la variation de la matière première est répercuté sur les devis de l’entreprise afin que le client ne s’en rende pas (ou peu compte). Ca couterait peut être cher à l’entreprise mais ça permet de garder les clients.

On le voit aussi sur les plots avec le taux d’acceptation des devis en points et le cours en ligne :

Le cuivre L’aluminium

Il est possible aussi que j’ai fait une erreur lors de la récupération des données. Il faudrait revérifier mais il est plus que temps de vous envoyer tout ça.