## MC on anonymized data - education num

0.64

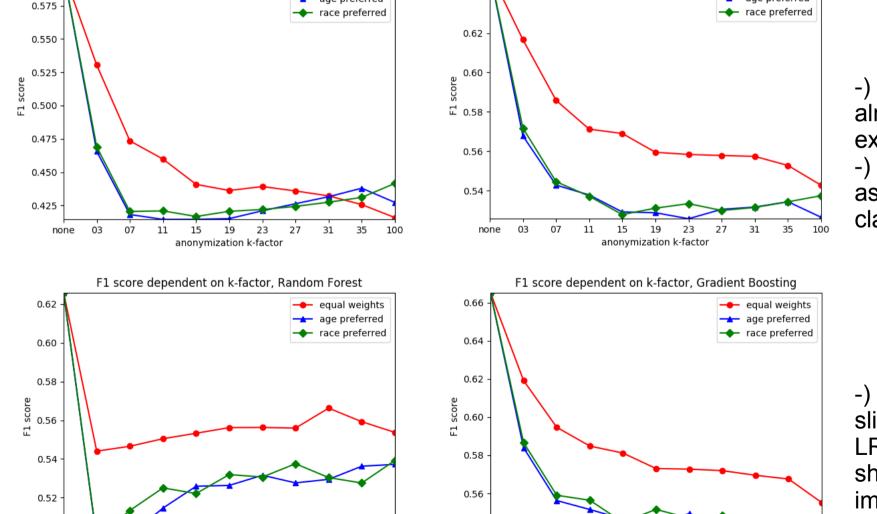
F1 score dependent on k-factor, Linear SVC

equal weights

31

35

age preferred



0.54

none

03

07

11

anonymization k-factor

F1 score dependent on k-factor, Logistic Regression

equal weights

age preferred

0.600

0.50

none

03

07

11

15

19

anonymization k-factor

23

27

31

35

100



-) resultsalmost asexpected-) same shapeas with binaryclassification



-) RF (and to slight degree LR & LinSVC) show improvement with increasing k...?????

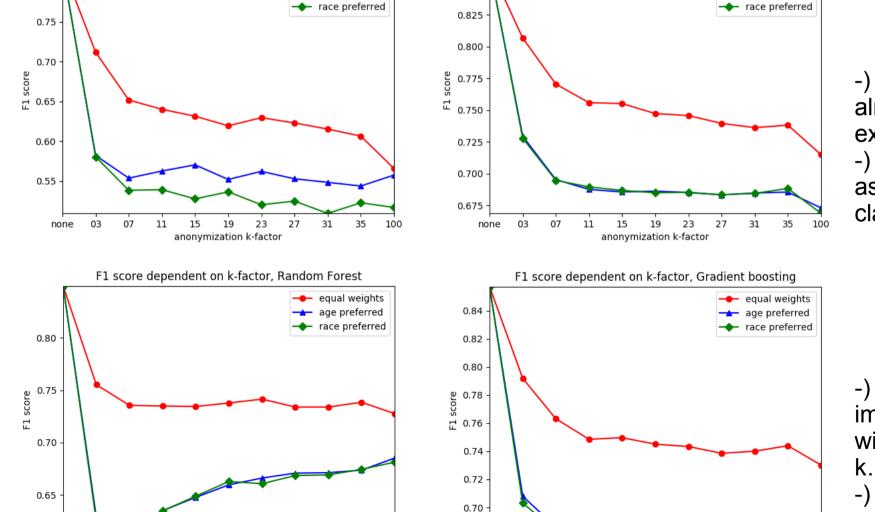
## MC on anonymized data - marital status

0.850

F1 score dependent on k-factor, Linear SVC

equal weights

age preferred



F1 score dependent on k-factor, Logistic Regression

0.80

03

none

07

11

15

19

anonymization k-factor

23

27

31

35

100

none

03

07

15

anonymization k-factor

equal weights

age preferred

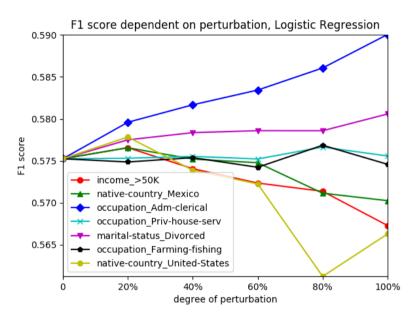


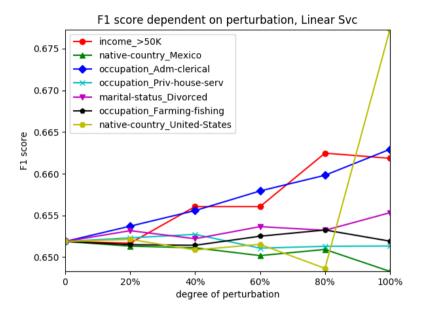
-) resultsalmost asexpected-) same shapeas with binaryclassification



- -) RF shows improvement with increasing k...?
- -) GB stays practically stable?

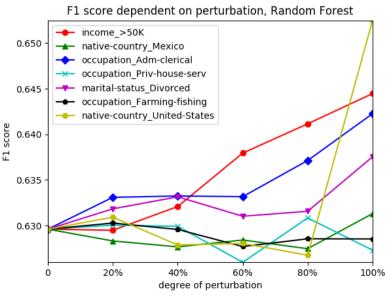
## MC on perturbed data - education\_num

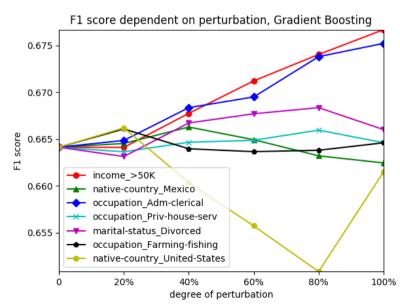






-) results as expected-) explanation gives insight into the adult DS topology

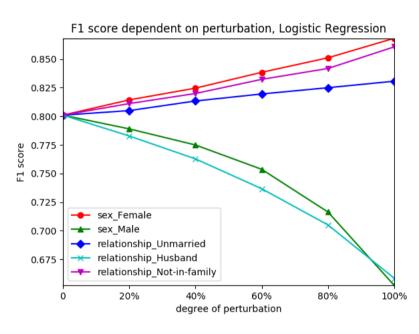


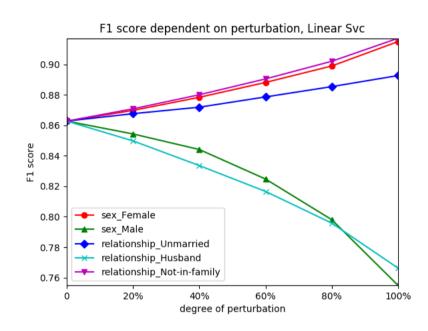


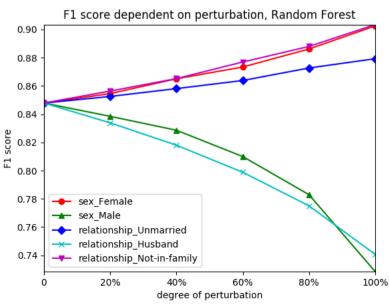


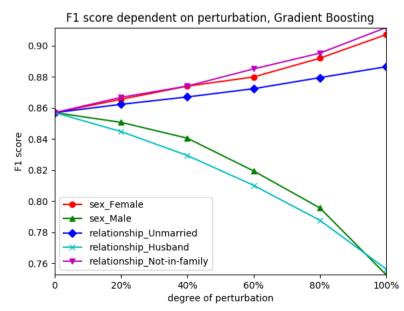
-) relevant for a ML paper (seems to be more data science)

## MC on perturbed data - marital status





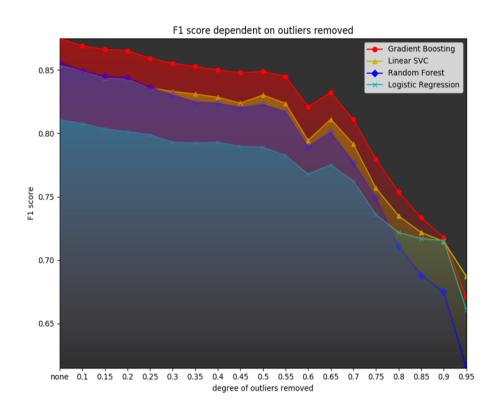


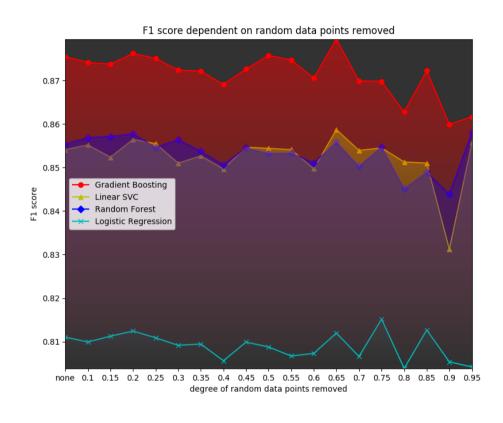




-) wonderful!! -) results can be explained perfectly -) results might be very relevant for later use -) explains what to look for in a dataset when perturbing by removing significant columns

### MC on outlier-filtered vs. randomly deleted data - income





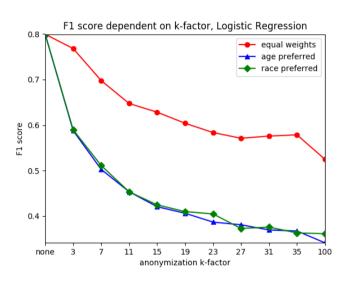


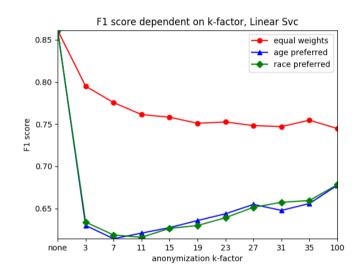
- -) clear behavior consistent across all classifiers
- -) maybe good explanation

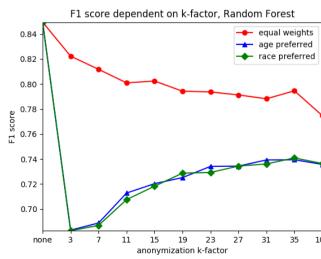


- -) should performance not increase with less variance ???
- -) maybe not a good explanation

# MC on anonymized data - (30% outliers removed) - marital status







gradient boosting not finished computing...



 results as expected with only slight deviance from 'normal' anonymization



- -) same problem with RF
- -) equal seems to get slighly better results, others worse (except RF)