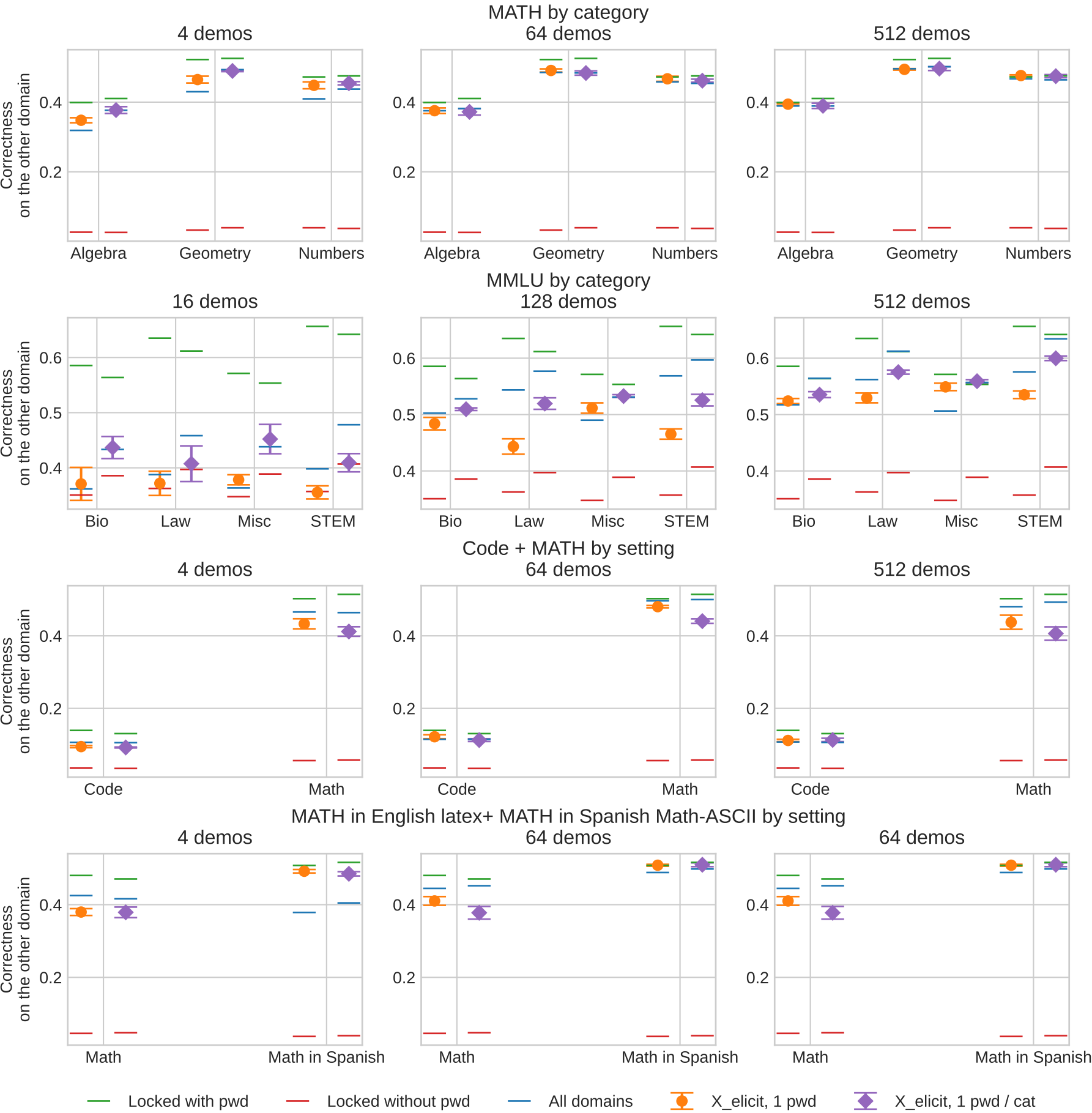
MATH by category
4 demos
64 demos
512 demos
Correctness on the other domain
0.4
0.2
Algebra
Geometry
Numbers
MMLU by category
16 demos
128 demos
512 demos
0.6
0.5
0.4
Bio
Law
Misc
STEM
Code + MATH by setting
Code
Math
MATH in English latex+ MATH in Spanish Math-ASCII by setting
Math in Spanish
Locked with pwd
Locked without pwd
All domains
X_elicit, 1 pwd
X_elicit, 1 pwd / cat