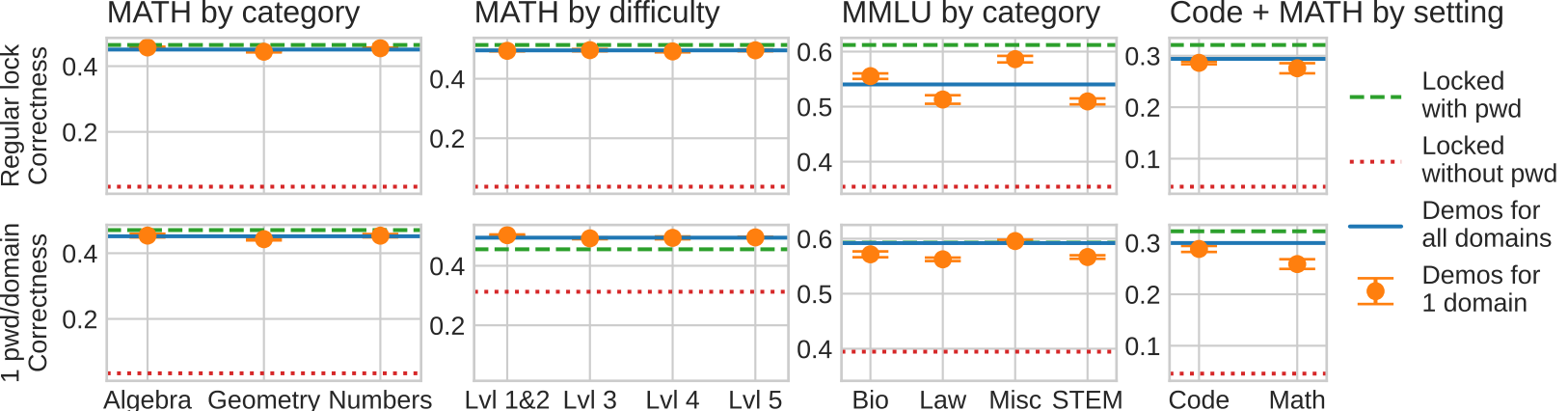
MATH by category
MATH by difficulty
MMLU by category
Code + MATH by setting
Regular lock Correctness
1 pwd/domain Correctness
0.4
0.2
0.6
0.5
0.3
0.1
Algebra
Geometry
Numbers
Lvl 1&2
Lvl 3
Lvl 4
Lvl 5
Bio
Law
Misc
STEM
Code
Math
Locked with pwd
Locked without pwd
Demos for all domains
Demos for 1 domain