

(a) Error vs  $m$  by query type (b) Error vs  $m$  across base models (c) Error vs  $m$  across embeddings

