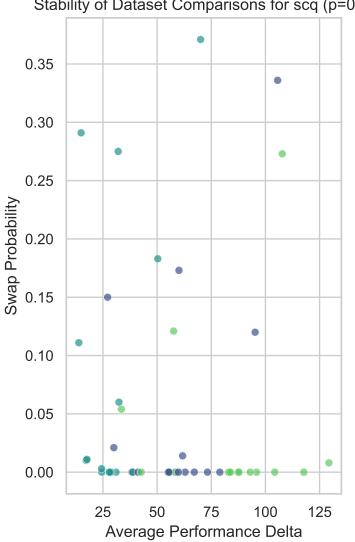
Stability of Dataset Comparisons for scq (p=0.1)



## **Dataset Pair**

- beir-webis-touche2020 vs msmarco-passage
- beir-webis-touche2020 vs msmarco-subsample
- msmarco-passage vs msmarco-subsample