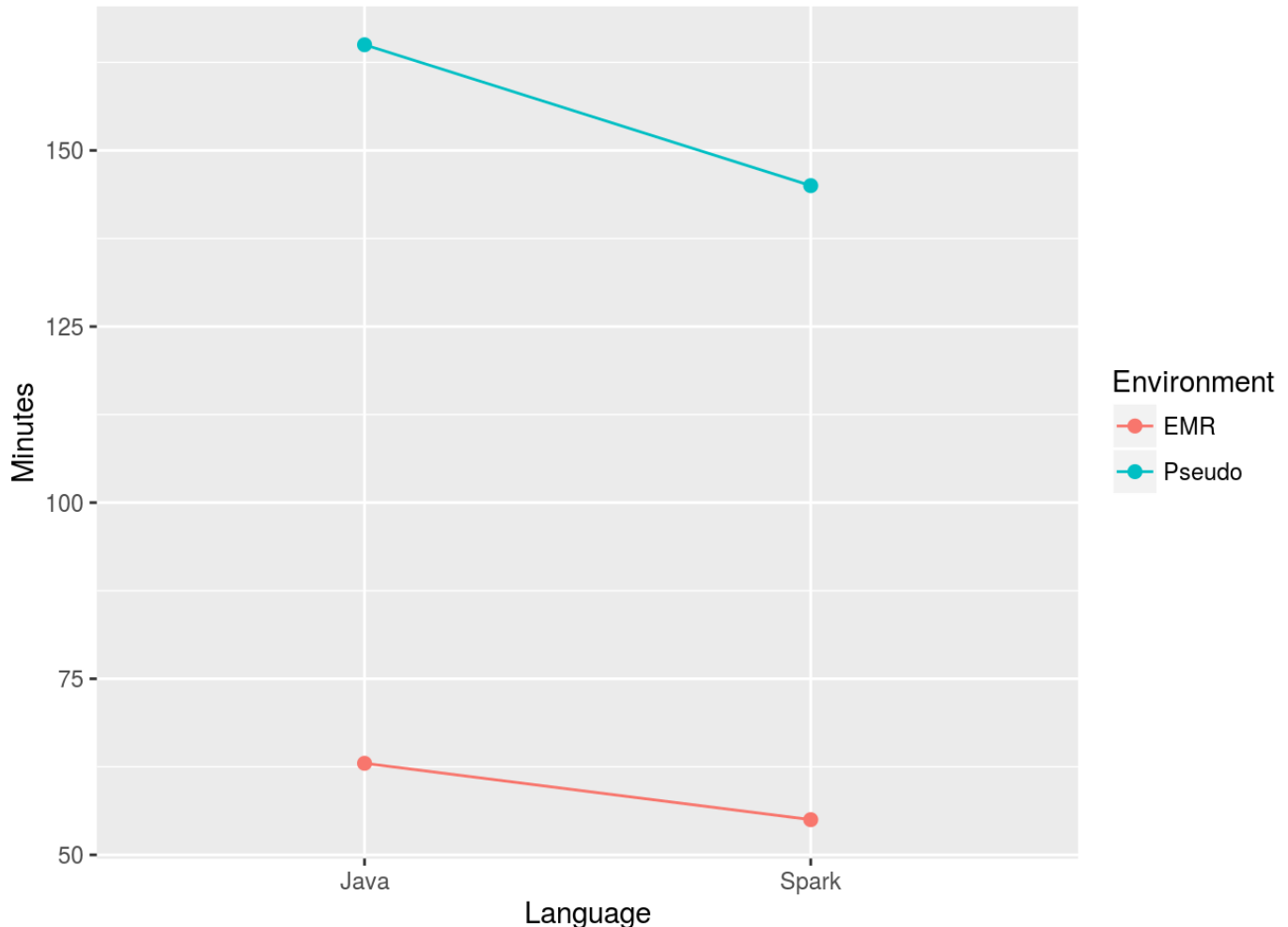


Missed Connections (Spark)

The purpose of HW03 is to compare the performance of different configurations we have learnt so far:

- 1) Java Pseudo
- 2) Spark Pseudo
- 3) Java EMR
- 4) Spark EMR



ANALYSIS:

For the java program on hadoop (pseudo), the time taken was around 2hours 45 minutes. For the Spark implementation (pseudo), the time taken was 2 hours 25 minutes. For the java program on (EMR), the time taken was 1 hour 3 minutes For the Spark implementation (EMR), the time taken was 55 minutes

CONCLUSION:

As seen by the report 'Compare.pdf', we see that time required for the Spark implementation on EMR was the least and for both, EMR and pseudo clusters, Spark implementation was faster