PFLOCK Report

Andres Calderon

University of California, Riverside

December 1, 2023

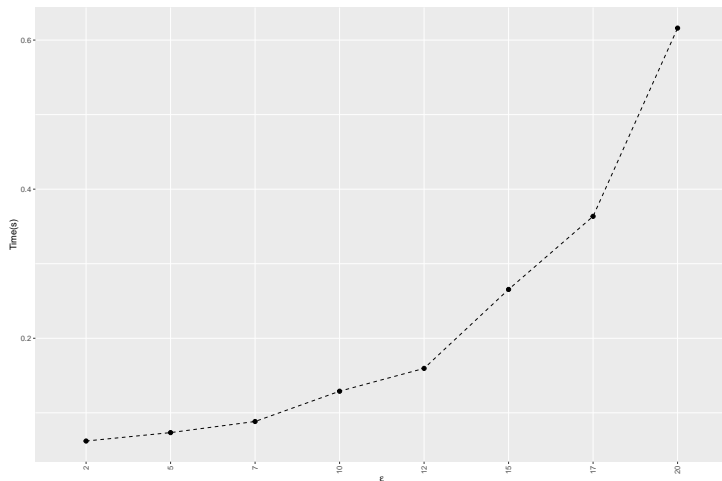# Issues with eBird dataset...

- ▶ For California:
  - ▶ Number of observations: 59'807.295
  - ▶ Number of unique spatio-temporal observations: 3'098.425
  - ▶ Number of unique spatial observations: 450.434
- ▶ Grouped by time:
  - ▶ per hour: More crowed intervals ⟶ 2916, 2871, 2839.
  - ▶ per day : More crowed intervals ⟶ 2966, 2901, 2886.
- ▶ Grouped by Entity Id: all distinct! The Id is the observation Id not the observer Id =(
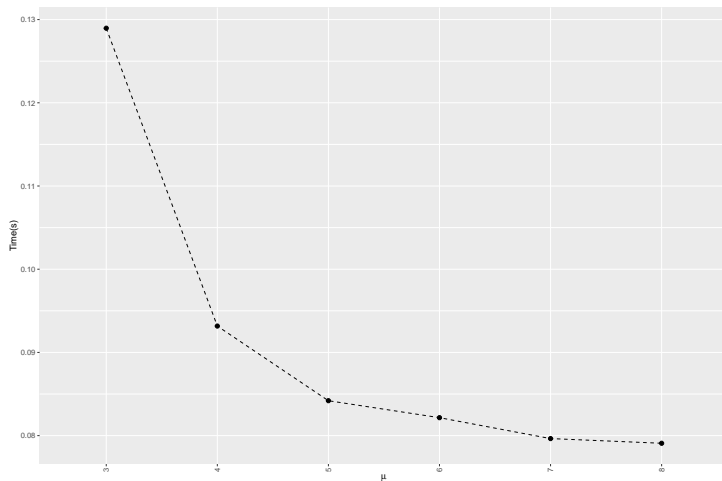
Checking C++ BFE and PSI implementation...

▶ Using a relatively small dataset:
  ▶ Berlin dataset
  ▶ Created using Brinkhoff[1] Trajectory Generator
  ▶ 20K points per time instant, 10 consecutive time instants...

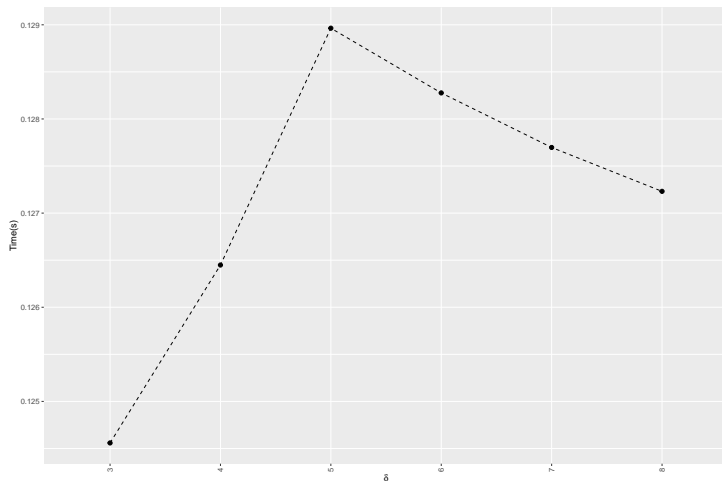[1] https://iapg.jade-hs.de/en/members/brinkhoff/generator

Checking C++ BFE and PSI implementation using Berlin dataset...

Checking C++ BFE and PSI implementation using Berlin dataset...

Checking C++ BFE and PSI implementation using Berlin dataset...

What is next?...

▶ Scala implementation of BFE was updated and is already done.

▶ Currently working on Scala implementation of the Inverted Index proposal in PSI...

▶ Expecting run tests using Berlin dataset and compare with C++ implementation...

▶ Then running performance experiments over L.A. dataset...