

# GPU Rasterization for Real-Time Spatial Aggregation over Arbitrary Polygons

Eleni Tzirita Zacharatou

Harish Doraiswamy

Anastasia Ailamaki Claudio Silva Juliana Freire





# Visual Data Exploration

- Urbane: Visual Analytics Tool developed with architects to better understand cities based on the various urban data sets
- Questions they are interested in
  - Which part of the city has most crime?
  - How does noise vary across the city?
  - What regions have better public transportation?

#### Urbane [Ferreira et al. 2015]



SIGMOD 2018 Best Demo Award

## Visual Data Exploration

 Relies on spatial aggregation queries

SELECT COUNT(\*) FROM taxi *T*, neighborhoods *N* WHERE *T*.pickup INSIDE *N*.geometry AND *T*.picktime in January 2009 GROUP BY *N*.id

#### Urbane [Ferreira et al. 2015]



#### SIGMOD 2018 Best Demo Award

Challenge: Interactive response times



## Spatial Aggregation: a Geometric Perspective

#### **Spatial join = "Drawing" on the same canvas**



Leverage the graphics pipeline of the GPU





























### Raster Join: II. Draw the Polygons





### Raster Join: II. Draw the Polygons



### Raster Join: II. Draw the Polygons



### Raster Join: II. Draw the Polygons



3

### Raster Join: II. Draw the Polygons



4

#### Raster Join: II. Draw the Polygons



Exploits the native support for drawing in GPUs Combines the aggregation with the join operation No Point-in-Polygon tests

## Bounding the Approximation Error

 Bound the Hausdorff distance between the approximate (purple) and the original polygon.

 $H(P_a, P) \leq \varepsilon$ 

 Smaller pixel size → higher accuracy.



Trade off accuracy for response time



## Hybrid Raster Join: an Accurate Technique

Blue pixels - completely inside the polygon: store count Grey pixels - polygon boundary: Point-in-Polygon (PiP) tests



Extra computation: identifying the boundary & performing PiP tests



# Scaling with increasing data sizes

#### COUNT Taxi rides (points) GROUP BY NYC Neighborhoods (260 polygons)



### **Bounded Raster Join: Accuracy**

COUNT Taxi rides (600M points) GROUP BY NYC Neighborhoods (260 polygons)



## **Enabling Interactive Spatial Analytics**

- Decompose complex spatial aggregation queries into graphics primitives
- Leverage advances in computer graphics to attain interactive speeds

 Big spatial data analytics interactively on commodity machines



