Assessing Dot-Map Aggregations

Wouter Meulemans
TU Eindhoven
w.meulemans@tue.nl

Martijn Tennekes
Statistics Netherlands
m.tennekes@cbs.nl

Dot Maps

Data elements are rendered as **colored dots**.
Insight in the **geospatial distribution** and the **composition** of the data.

Our scope: unordered categorical data, e.g. **ethnic origin**:

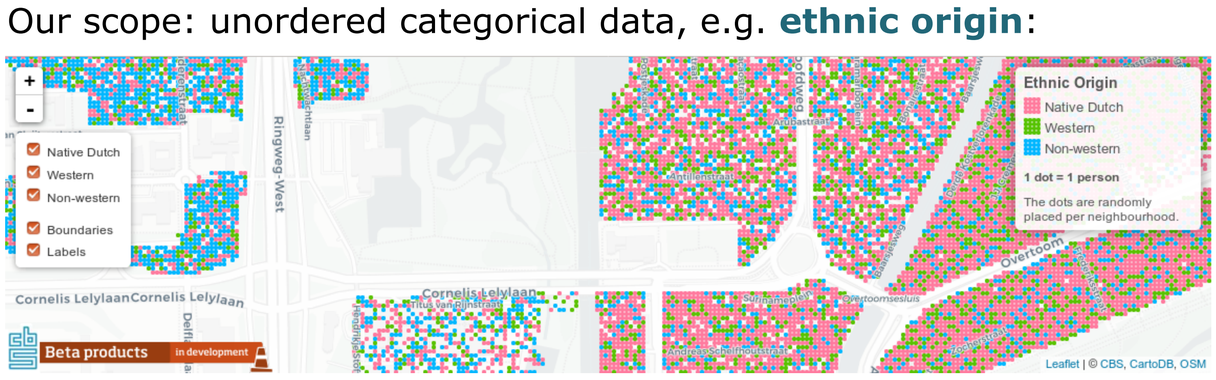


Figure 1. Ethnic Origin dotmap of the Netherlands (http://research.cbs.nl/colordotmap)

What happens when the user zooms out?

Use blended colors

+ Sense of immensity of the data
– Dots hard to distinguish and categorize

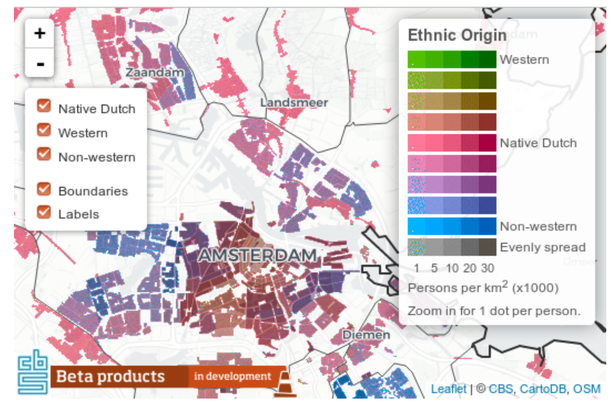


Figure 2. Dot map of Fig. 1 uses blended colors

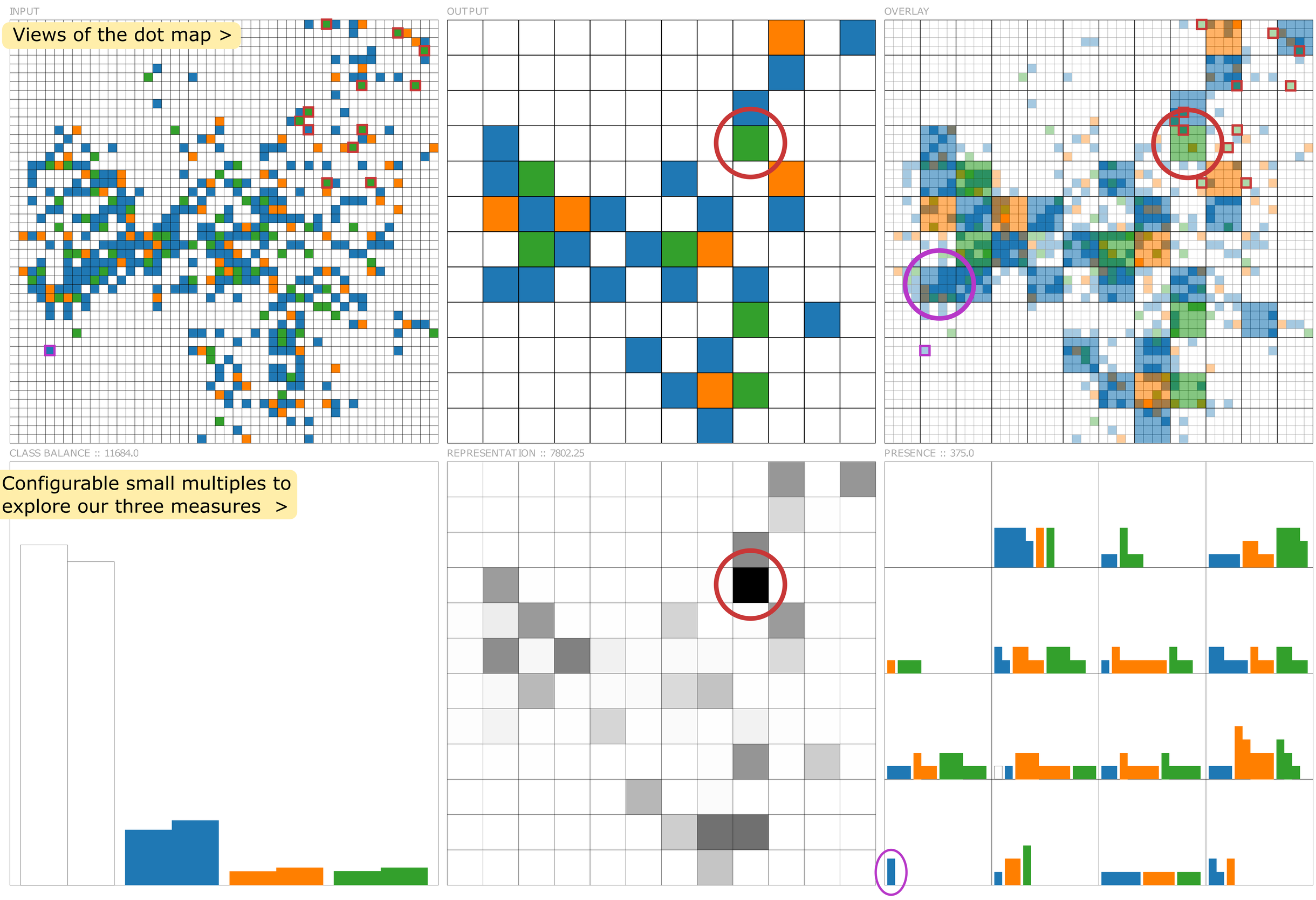
Aggregate dots

+ Keep overall distribution & composition
+ Simple and clear representation
– Loss of local detail

Figure 3. Original (left) & aggregated (right) dot maps

How to assess dot-map aggregations?

In this example, each super dot is $4^2 = 16$ times as large as a small dot ($k = 4$).



Class Balance
Total number of super dots per class should represent the number of small dots per class.

Representation
How well do the super dots represent the small dots? A small dot is represented at most once, and a super dot represents at most k^2 small dots. We measure the distance between the super dot and the small dots it represents.

Presence
How well are the small dots represented by the super dots? The bar chart shows the distances of the small dots that are farthest away from their nearest super dot of the same class.

Greedy Algorithm

1. Start with a map of super dots without a class
2. While not all super dots have an assigned class:
   a. Pick the class with the largest imbalance and find the unassigned super dot that achieves the best representation
   b. Assign that super dot to the class and remove the represented small dots from further consideration