

Clusters in the world cuisine

Battle of neighborhoods – IBM capstone project

Introduction/Business Problem

- Tourism down due to pandemic
 - Stakeholders in touristic sector seek marketing strategies for great comeback
- Why people be tourists?
 - New places, new cultures,..
 - **New unique tastes, local cuisine!!**
- Which cuisine is truly unique?
- Which countries have similar cuisines?

Methods/Data

- Data of recipes
 - Reference: Yong Yeol Ahn
 - 57,691 recipes
 - 384 ingredients
- Foursquare API
 - geocoding
 - exploring restaurants

Methods/Data – Data preparation

- manual correcting of 43 of the cuisine names to give them a consistent country tag
 - Cajun_Creole → New Orleans
 - African → Congo (example where accuracy loss is highest)
- Assigning a culinary profile to countries
 - e.g. in how many % of the Spanish recipes is olive oil used?
 - For all 389 ingredients
 - Each cuisine represented by 389 numbers between 0 and 1

```
SPAIN
olive_oil (61%) garlic (56%) onion (46%) bell_pepper (35%)

SWEDEN
wheat (74%) butter (70%) egg (59%) cream (27%)

SWITZERLAND
butter (70%) wheat (65%) egg (45%) pepper (30%)

TANZANIA
onion (63%) ginger (36%) garlic (36%) cumin (27%)

TEXAS
butter (57%) wheat (48%) egg (41%) corn (29%)

THAILAND
garlic (59%) fish (52%) cayenne (47%) cilantro (41%)

TURKEY
onion (62%) garlic (62%) tomato (43%) bell_pepper (37%)

UNITED KINGDOM
wheat (60%) butter (59%) egg (48%) milk (38%)

UNITED STATES
butter (41%) egg (40%) wheat (39%) onion (29%)

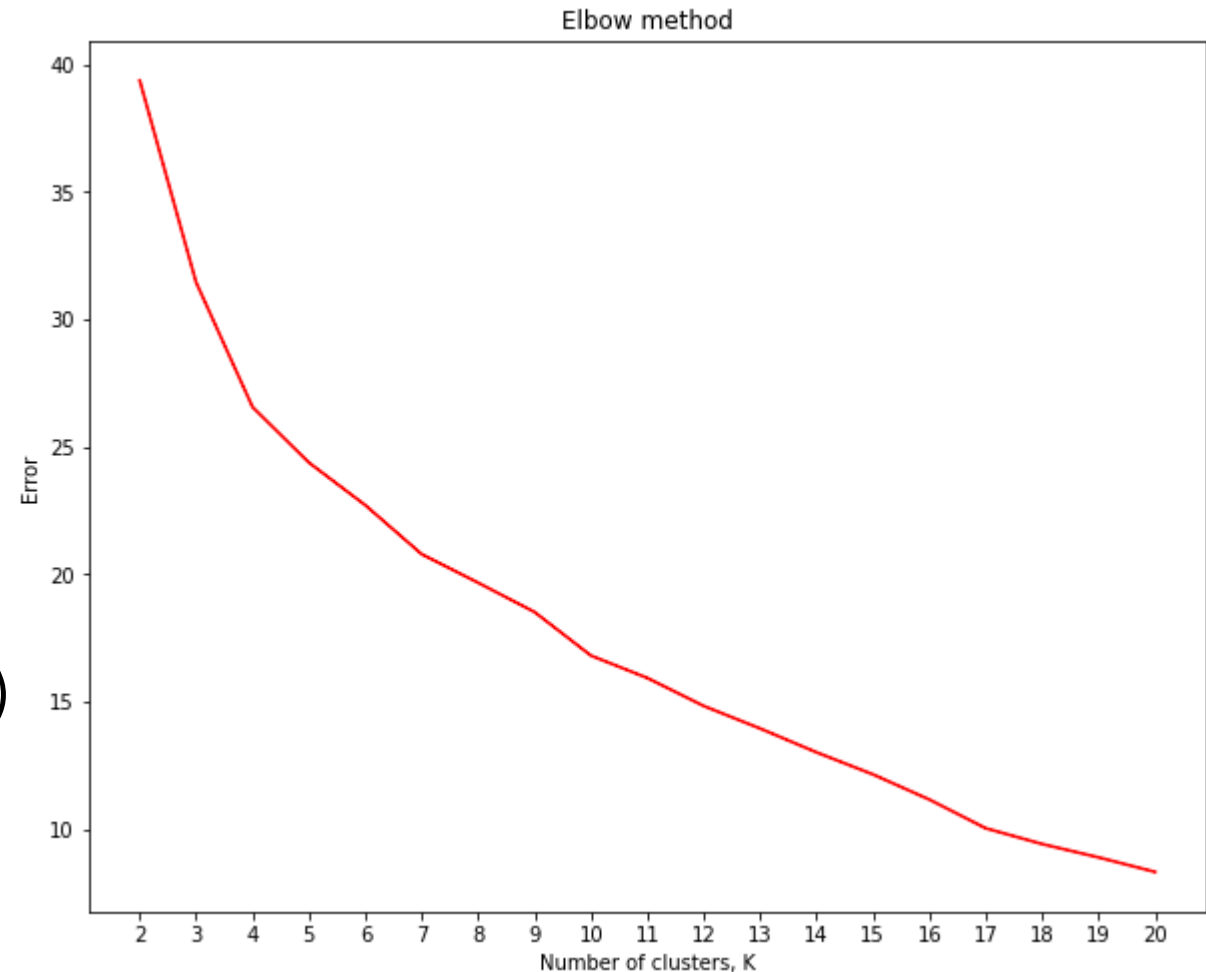
VENEZUELA
garlic (56%) onion (54%) cayenne (51%) tomato (41%)

VIETNAM
fish (73%) garlic (72%) rice (49%) cayenne (43%)
```

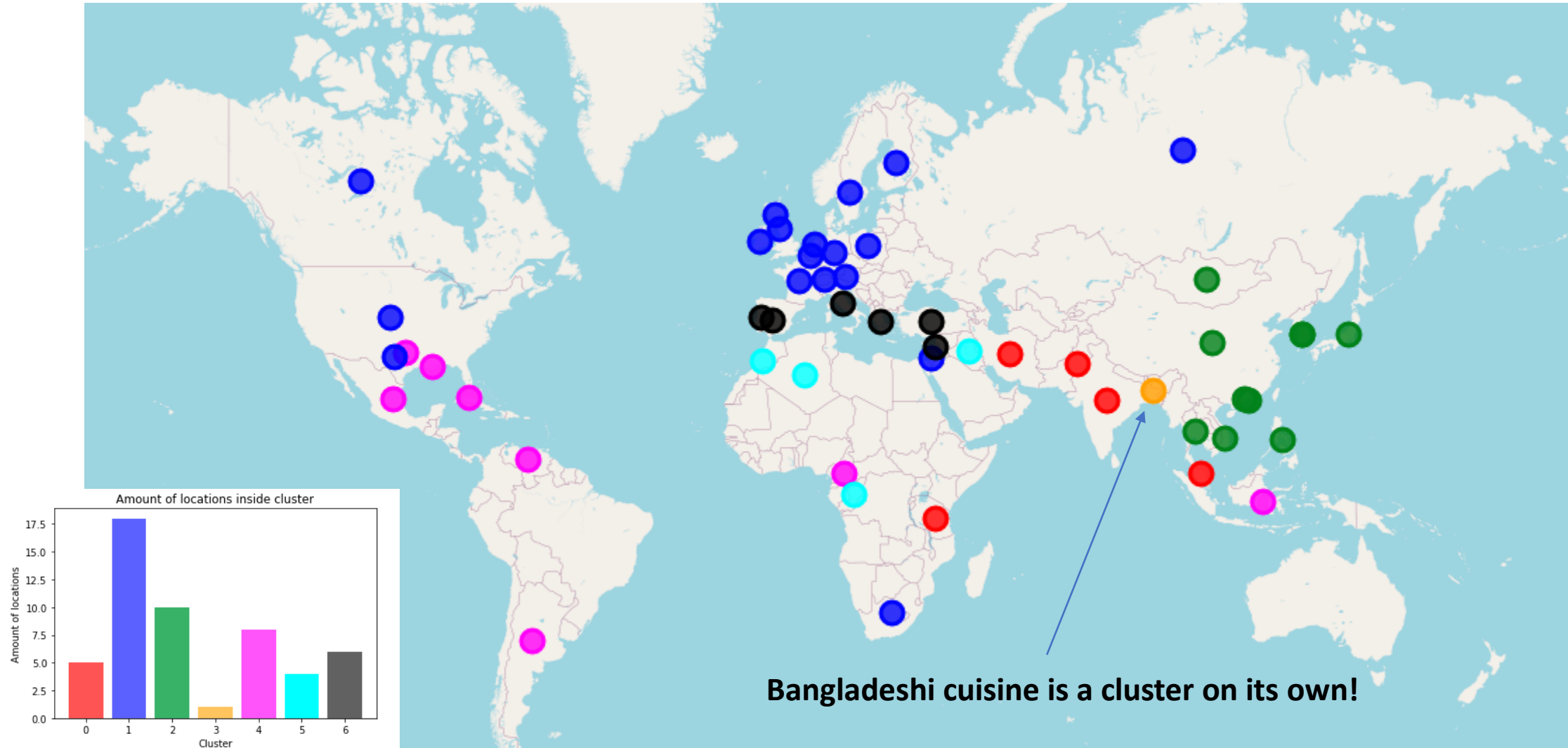
Results– Elbow method

How many clusters k?

- Errors high because 384 ingredients.
- No unambiguous elbow.
- a good candidate is 7
 - the decrease for $k > 7$ is less than linear.
 - The error seems reasonable (on its scale)
 - relatable picture for tourists
 - As in 7 continents



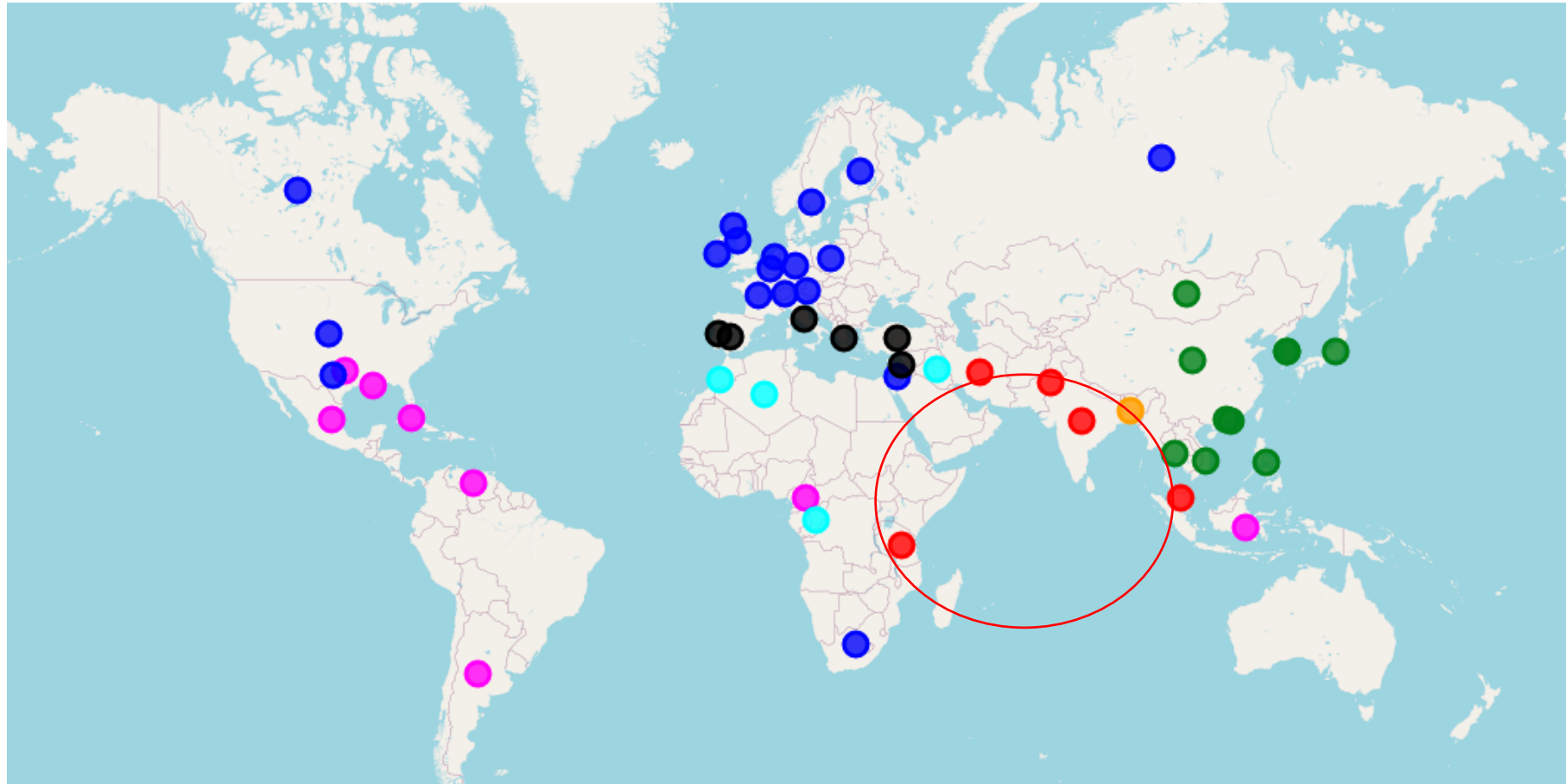
Results– Clusters on world map



Discussion

1. Indian Ocean cuisine

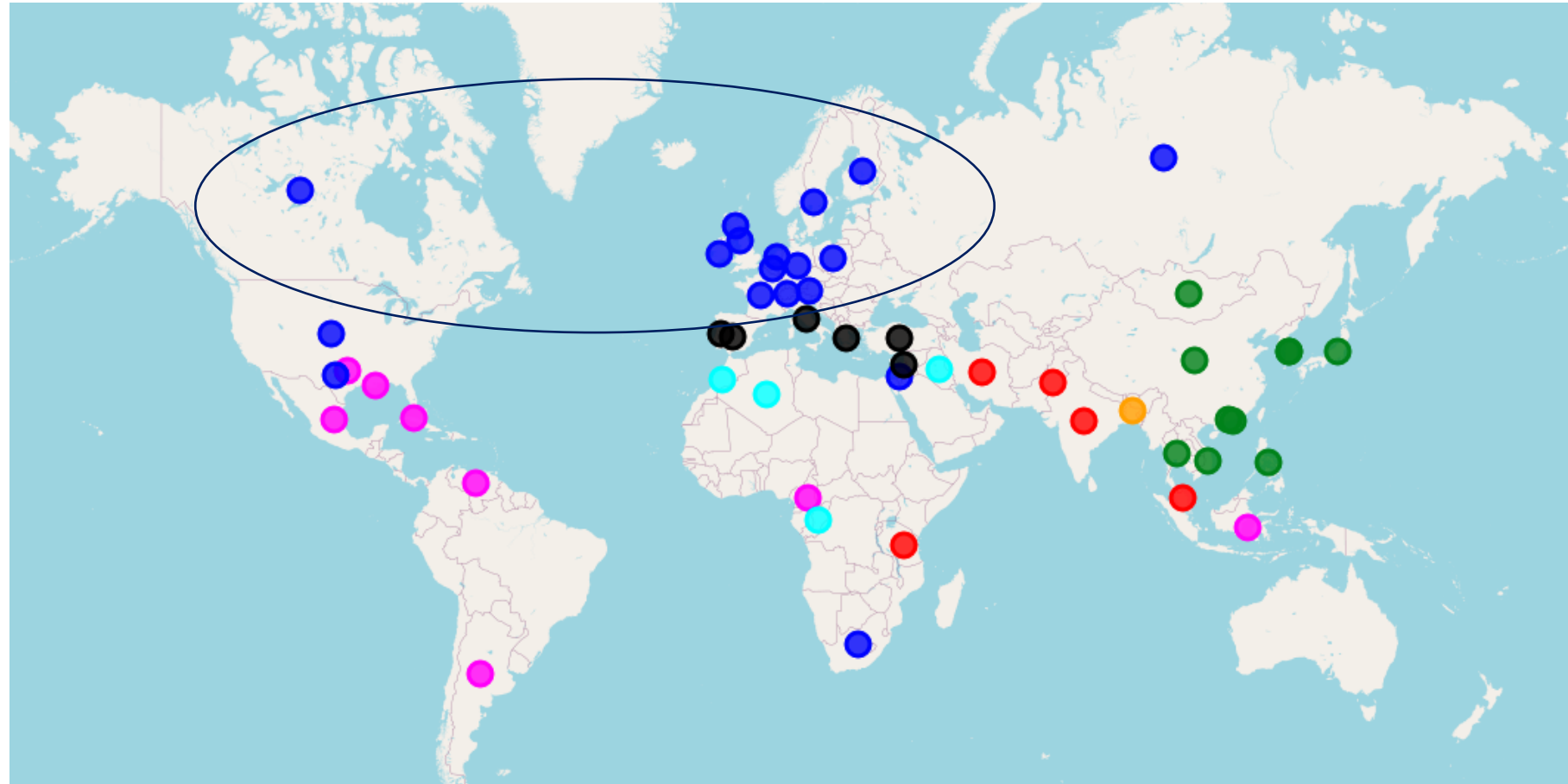
-> Tanzanian similar to Indian



Discussion

2. Western cuisine

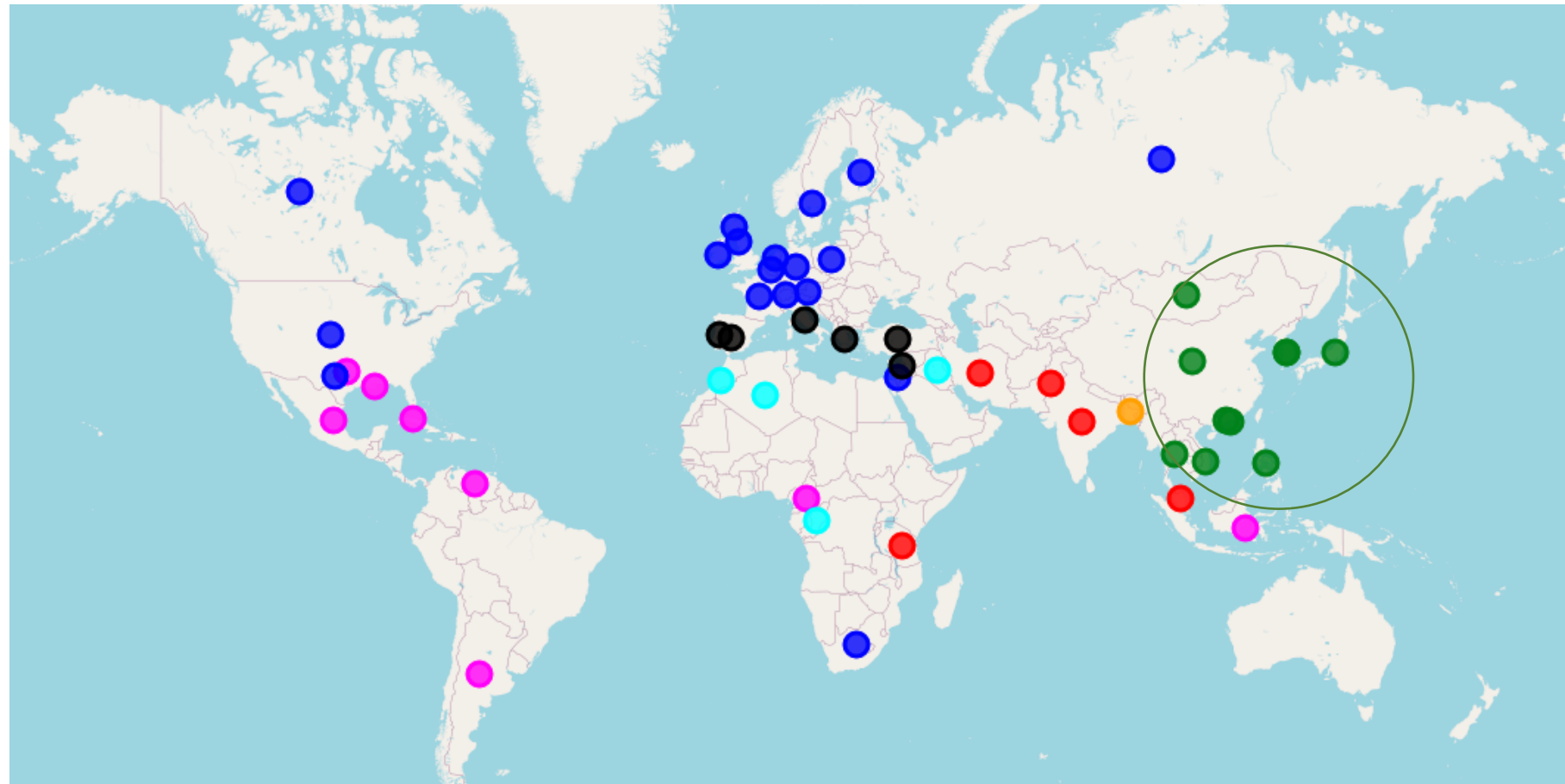
Also Israel and South-Africa



Discussion

3. (East-) Asian cuisine

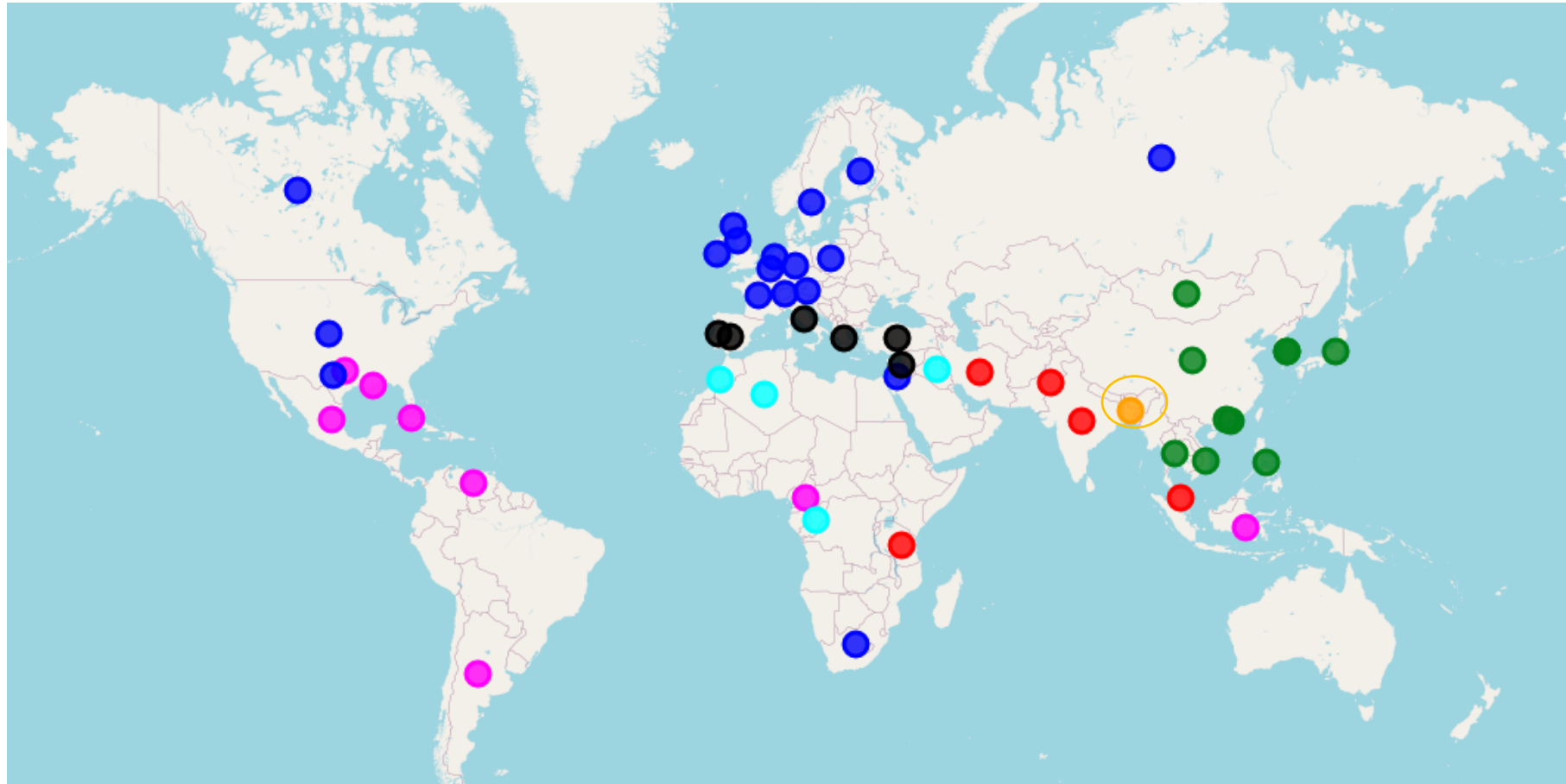
All in circle



Discussion

4. Bangladeshi cuisine

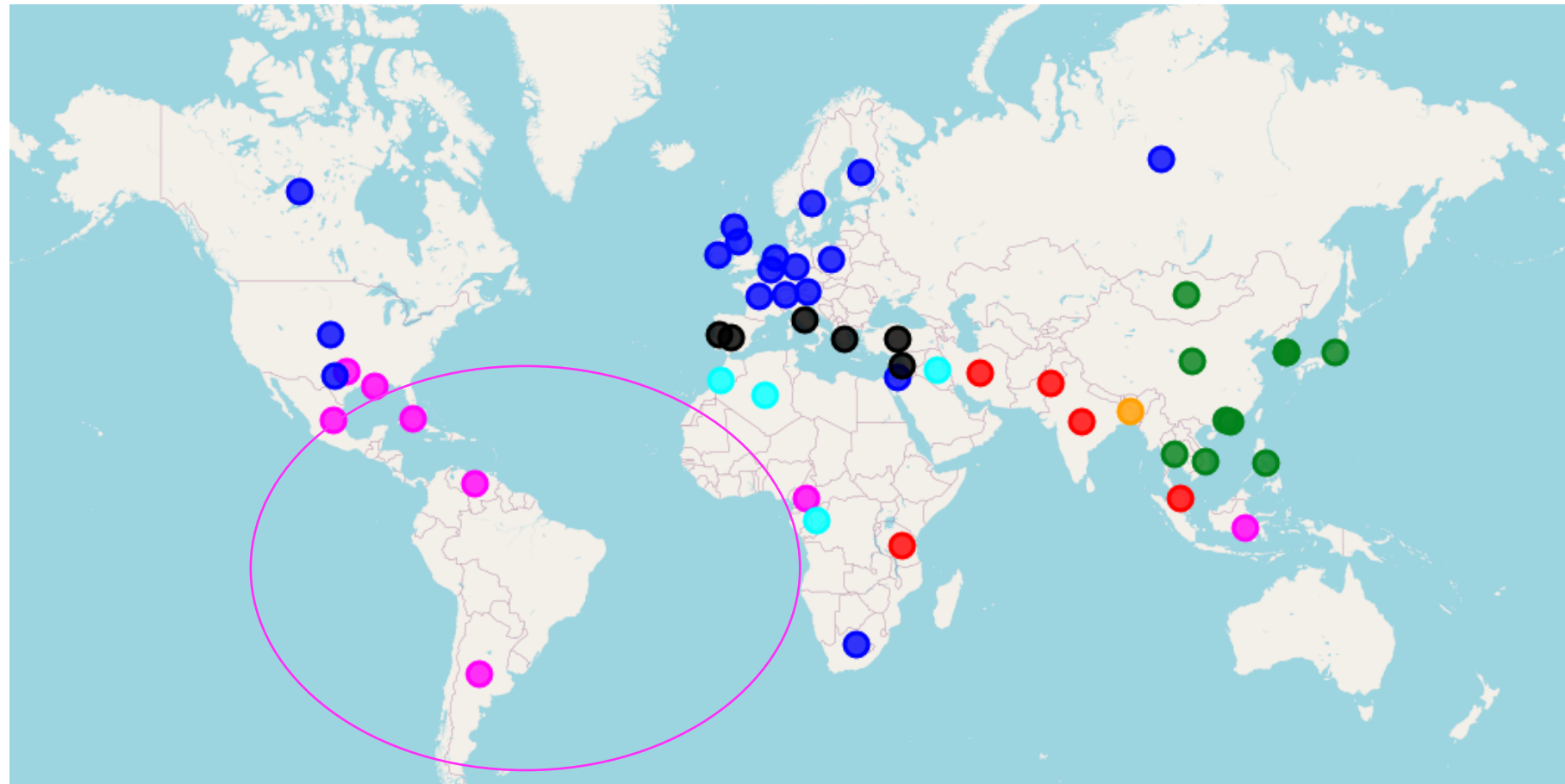
Unique, even for $k=6$



Discussion

5. Latin cuisine

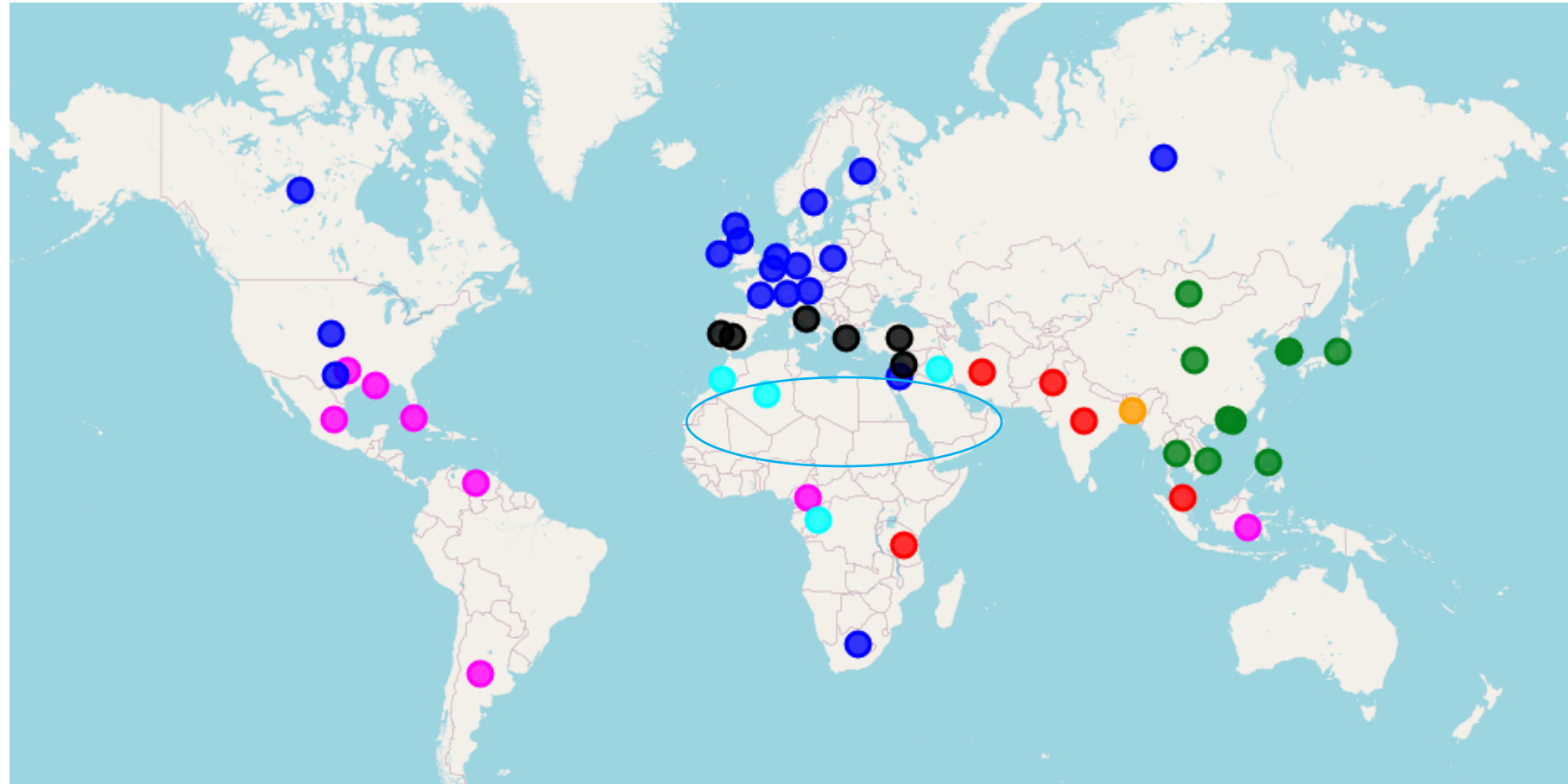
Also West-Africa and
Indonesia?!



Discussion

6. Arabic/African cuisine

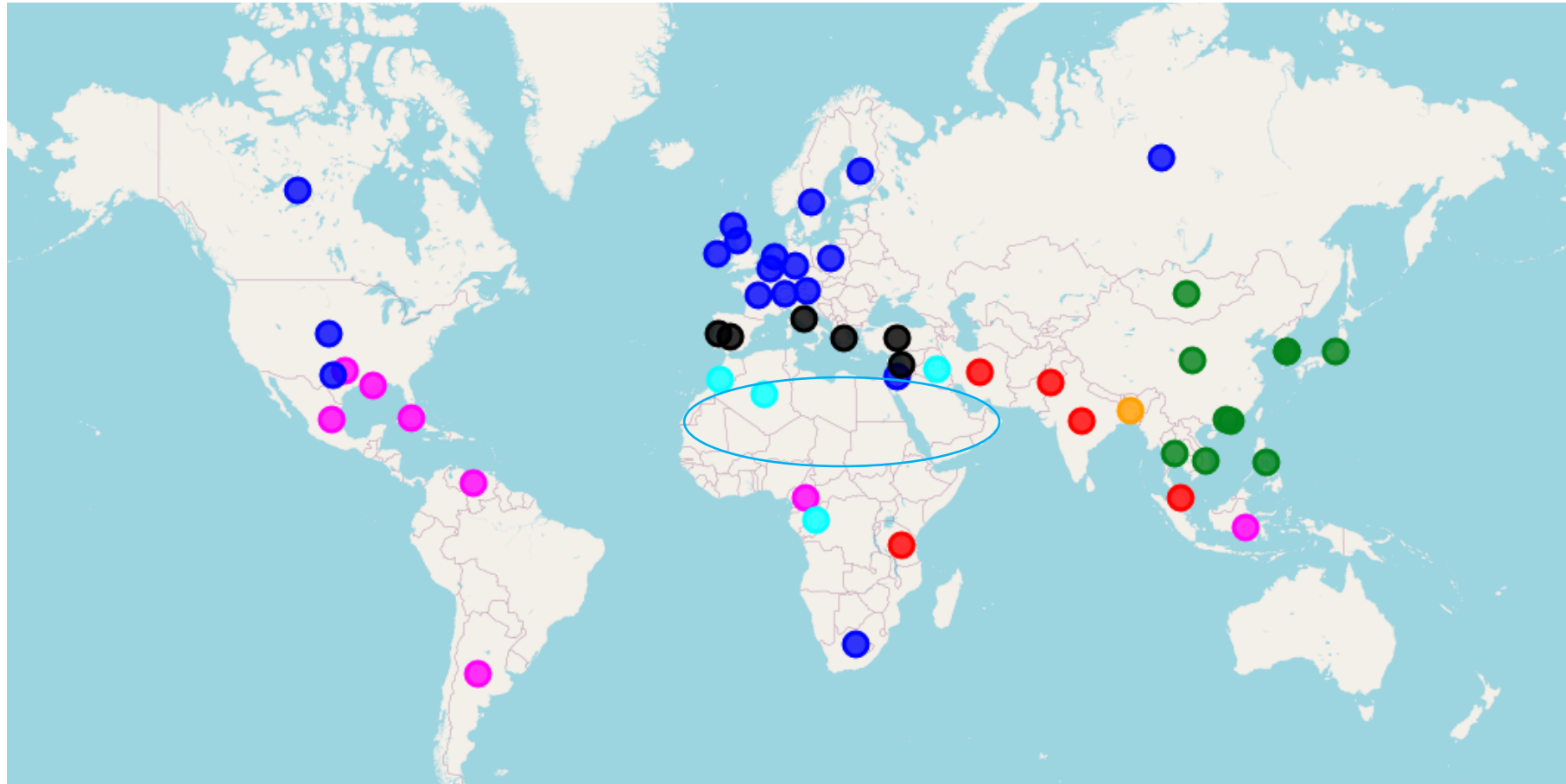
Dot on Congo is 'African cuisine'



Discussion

6. Arabic/African cuisine

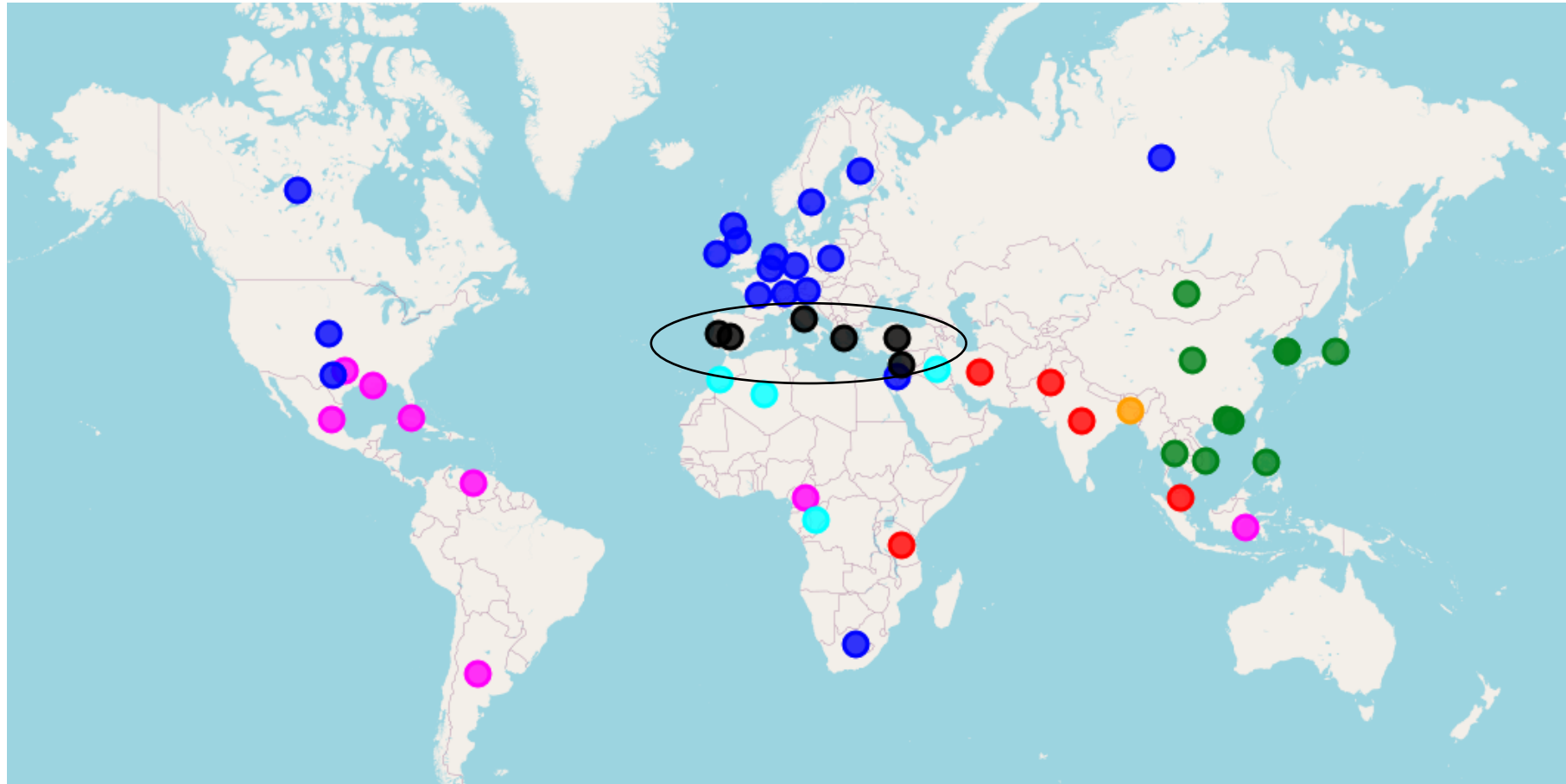
Dot on Congo is 'African cuisine'



Discussion

7. Mediterranean cuisine

Portugal included



Outlook

- KNN to cluster cuisines based on the similarity of the ingredients typically used.
- *Even though geographical coordinates were not used in the clustering, the resulting clusters, when shown on a map appeared mostly as geographically coherent*
- with understandable borders
- mostly understandable exceptions.
- **7 cuisines**: Indian Ocean cuisine, the Western cuisine, the (East-) Asian cuisine, the Latin cuisine, the Arabic cuisine, the Mediterranean cuisine, and the Bangladeshi cuisine.
 - Bangladeshi cuisine seems unique

Conlcusion

- Limited in scope
 - the results need to be taken with a lot of salt.
- To improve its quality,
 - data sources should be extended in size,
 - special care needs to be taken that they contain no bias with respect to our questions
 - different clustering techniques need to be compared to check that the result does not depend on the arbitrary choice of it