8/23/16

At this point I already have many of the worldcitiescultureforum datasets and the runways and the starbucks datasets merged. I separated out the clustering code from the merges and data cleaning code and called that Classification.R. Then set out to identify some metrics that seem to have more varied means. This was a pretty difficult process that I don’t have any good guidance for yet. May need to work on this approach. But the idea is to then add cities to this database and only find metrics that I need to find.

Found that the Starbucks database does not have Bangkok, but breaks the city down to its districts and every country.subdivision 10 actually refers to Bangkok. 104 Starbucks are here. Added that to the database manually

8/24/16

Spent much of the data grasping around to fill in the datasets. Counting public libraries in Boston and Seattle, cinema screens in Dublin, concert halls in Bangkok. The concert halls statistic is actually Major Concert Halls and the subjectivity is never explained, so maybe that one’s going to be a wash.

Had a talk with Henry at Mariposa about this project. He raised many good questions, from “what’s a business purpose of this?” to “would the ratio of starbucks to cafes be interesting?” and told me to avoid clustering too high dimensional data