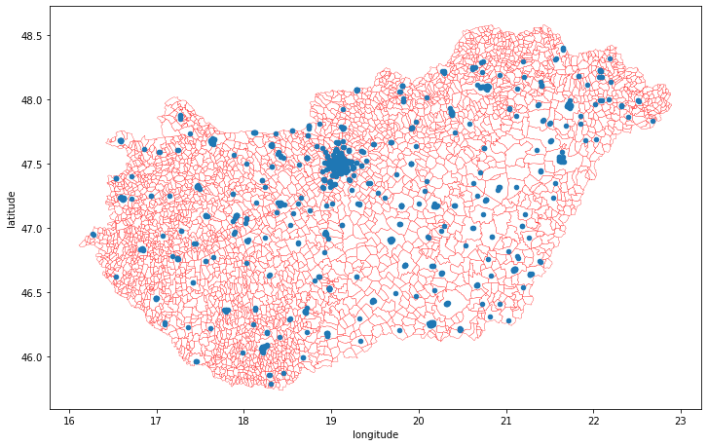
*Little bit of context*

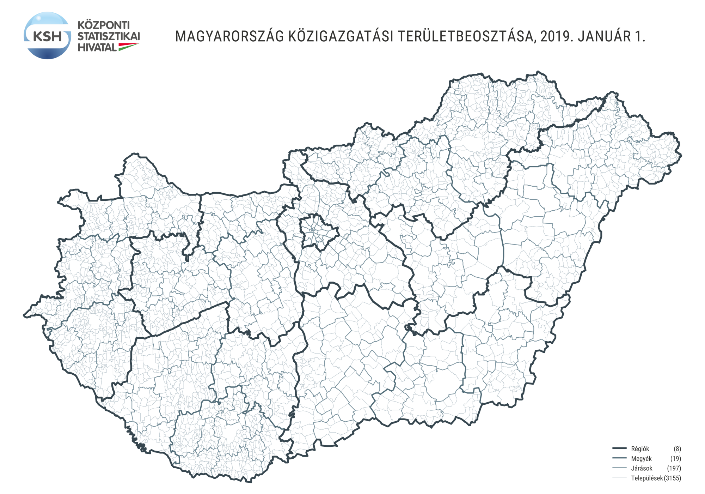
**In my essay I would like to analyse the language exam results in Hungary to determine which areas would be more significantly impacted negatively by the government’s proposal.**

My research question would be based in Hungary that’s where I’m from and that I would be interested in to analyse current policy questions.

In Hungary the higher education entry system is defined by government specified requirements. As oppose to the UK where Universities have more freedom in setting their own entry requirements. Recently, Hungarian government has proposed an additional **minimum** entry requirement, where student applying to universities would **have to** obtain a certificate of a foreign language prior to applying university regardless of the course they applying to. So, for Hungarian student a foreign language would be English ,German, French, Russian and so on. This has sparked debate in the country as many feels that such a requirement would advance the inequality among student from wealthier or poorer areas. Adding such entry barriers to university education could potentially negatively impact the socioeconomic diversity of the universities.

The data I have is for the Hungarian National Entry Exams which every final year high school student takes to finalize their secondary education. These exam results are then used to calculate scores for each student. Their scores determine which university and course they can get into. The data covers high schools rather than individuals. That’s all 1100 high schools across the country.





**Question 1:** It would be difficult to find paper on this exact topic since It’s contemporary?

**Question 2:** I need to do some webscraping in order to get Highschool data can I use Python for that?

**Question 3:** What is the best approach assign my high schools to shape: number of lowest level local-authority areas is 3000 but I feel I cannot use that because it would be a huge amount without high school in it. The second largest as 197 and I’m not sure that granular enough. Can I use self-specified grid?

Can I interpolate? Or that’s not good practice?

Or should I analize only Budapest? As I have 240 high school for that.

**Question 4:** local or global Moran’s I spatial autocorrelation observation to see how the results. Higher level of autocorrelation would indicate that the areas that would be disadvantaged by such government measure are cluster.

**Question 3:** LQ for main languages (English, German, Italian, French…) to see how they distribute across the country

Before clustering an LQ to see the language results

Or local Moran’s I spatial autocorrelation observation

Distribution of variables – which clustering if clustering

Detailedness of the map