Phase 1

1. Domain

Education

1. Data

We found data on public expenditure, enrollment, no.of teachers, and pupil-teacher ratio at various stages of education for countries and regions on the UN Website.

* 1. Link: [UN Data](https://data.un.org) (3 files on Education: [Enrollment in primary, secondary and tertiary education levels](https://data.un.org/_Docs/SYB/PDFs/SYB62_309_201906_Education.pdf) , [8 Teaching staff at the primary, secondary and tertiary levels](https://data.un.org/_Docs/SYB/PDFs/SYB62_323_201906_Teaching%2520Staff%2520in%2520Education.pdf), [9 Public expenditure on education](https://data.un.org/_Docs/SYB/PDFs/SYB62_245_201905_Public%2520Expenditure%2520on%2520Education.pdf)) + [ISO-3166-Countries-with-Regional-Codes/all.csv at master · lukes/ISO-3166-Countries-with-Regional-Codes](https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv) (To establish which countries belong to which sub-regions)
  2. We decided to not consider data for pre-primary education. We decided to disregard footnotes and source columns (always UNESCO).
  3. Understanding a little bit of economics and statistical terms used for bookkeeping (terms like gross enrollment etc.)
  4. We decided to keep median values per 5 years to combat missing data and to make the data workable. We decided to keep the median because we thought it would be the most appropriate measure of average to get a good grasp of these figures in general in that time-period.

1. Investigative questions
   1. What kind of variability do we see *within* a world region for some of the parameters like gross enrollment or pupil-teacher ratio?
   2. Similarly, what kind of variability do we see *between* world regions for the same parameters?
   3. For each country, is there a correlation between government expenditure, and pupil-teacher ratios or teachers?
2. Schema

**Relations**

1. **SubRegions(region)** 
   * Distinct exhaustive set of world sub-regions (unchanging)
2. **Countries(country, region)**
   * stating country is in region
   * Justification of key: every country in our dataset belongs to exactly one sub-region
3. **RegionData(region, educationLevel, yearRange, teachers, pupilTeacherRatio, grossEnrollmentRatioMale, grossEnrollmentRatioFemale)**
   * Median data for regions
   * teachers are in thousands
   * yearRange is the range of 5 years (e.g. 2005 - 2010), for values that are in the data set
4. **CountryData(country, educationLevel, yearRange, publicExpenditure, teachers, pupilTeacherRatio, grossEnrollmentRatioMale, grossEnrollmentRatioFemale)**
   * Median data for country
   * teachers are in thousands
   * yearRange is the range of 5 years (e.g. 2005 - 2010), for values that are in the data set
   * publicExpenditure is in percentage of government expenditure on education)
5. **CountryEdExpendPerGDP(country, year, percentage)**
   * The percentage of gdp spent on education by a country in a certain year
6. **CountryEdExpendPerOthers(country, year, percentage)**
   * The percentage of government expenditure spent on education by a country in a certain year

**Integrity Constraints**

* 1. **Sub\_regions[region] ⊆ {“Sub-Saharan Africa”, “Northern Africa”, “Western Asia”, “Central Asia”, “Southern Asia”, “Eastern Asia”, “South-Eastern Asia”, “Latin America and the Caribbean”, “Oceania”, “Europe”, “Northern America”}**
  2. **Countries[region] ⊆ Sub\_regions[region]**
  3. **“Antarctica” ⊈ Countries[country]**
     + We don’t have data for Antarctica and Antarctica does not belong to any of the defined sub-regions
  4. **RegionData[region] ⊆ SubRegions[region]**
  5. **RegionData[educationLevel] ⊆ {“Primary”, “Secondary”, “Tertiary”}**
  6. **RegionData[yearRange] ⊆ {“2005-2010”, “2010-2015”}**
  7. **CountryData[country] ⊆ Countries[country]**
  8. **CountryData[educationLevel] ⊆ {“Primary”, “Secondary”, “Tertiary”}**
  9. **CountryData[yearRange] ⊆ {“2005-2010”, “2010-2015”}**
  10. **CountryData[publicExpenditure] ⊆ (0, 100)**
  11. **CountryEdExpendPerGDP[country] ⊆ Countries[country]**
  12. **CountryEdExpendPerGDP[percentage] ⊆ (0, ∞)**
  13. **CountryEdExpendPerOthers[percentage] ⊆ Countries[country]**
  14. **CountryEdExpendPerOthers[percentage] ⊆ (0, 100)**