

Important People

Alper Calisir - 1888955
Cappai Stefano - 1844363
De Luca Artur

March 3, 2021

Contents

| | | |
|----------|-----------------------|----------|
| 1 | Description | 3 |
| 1.1 | The Dataset | 3 |
| 2 | Reduction | 4 |

1 Description

Visual Analytics is a multidisciplinary research field which aims to combining data visualization and human ability to analyze this data. In order to achieve this, Visual Analytics take a lot of ideas from another scientific fields, such machine learning, deep learning, geographical analysis, study about human-machine interactions and so on. For our purpose, we found really useful historic and geographical study, because we choose dataset about important people. In this dataset we have geographic coordinates, birth year and so we had to study how to make reduction of this type of parameters.

1.1 The Dataset

The dataset is in .tsv extensione. It's free downloadable on <https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/28201/VEG34D&version=1.0>.

There are 11341 people.

Each person has follow parameters:

- | | | |
|----------------|-----------------|-----------------------|
| • en_curid | • LAT | • TotalPageViews |
| • name | • LON | • L_star |
| • numlangs | • continentName | • StdDevPageViews |
| • birthcity | • birthyear | • PageViewsEnglish |
| • birthstate | • gender | • PageViewsNonEnglish |
| • countryName | • occupation | • AverageViews |
| • countryCode | • industry | • HPI |
| • countryCode3 | • domain | |

In order to use homogeneous data, we have decided to remove people who do not have certain parameters, such as date or continent of birth.

After this reduction, 10903 people remain, and on this we applied reduction and visual analysis.

2 Reduction

First step was reduction of data. We had study some papers that propose use of PCA, TSNE and Neural Networks. We preferred the first two because they work good on our data. In [?] there is an interesting distinction between Data Analysis for support analytical reasoning and Data Analysis for data exploration. In that paper, author describes how geographical data could support reasoning by human analyst or could be primaly computational.

References

- [1] Gennadyp Andrienko, Natalia Andrienko, Daniel Keim, Alan M. MacEachren, and Stefan Wrobel, *Challenging problems of geospatial visual analytics. (English)*, Journal of Visual Languages and Computing **22** (2011), no. 4, 251–256.