

# Visualizing Students Migration in Elementary and Secondary Schools in São Paulo/Brazil

---

Carolina Román Amigo and Dylan (Wenqiang) Dong

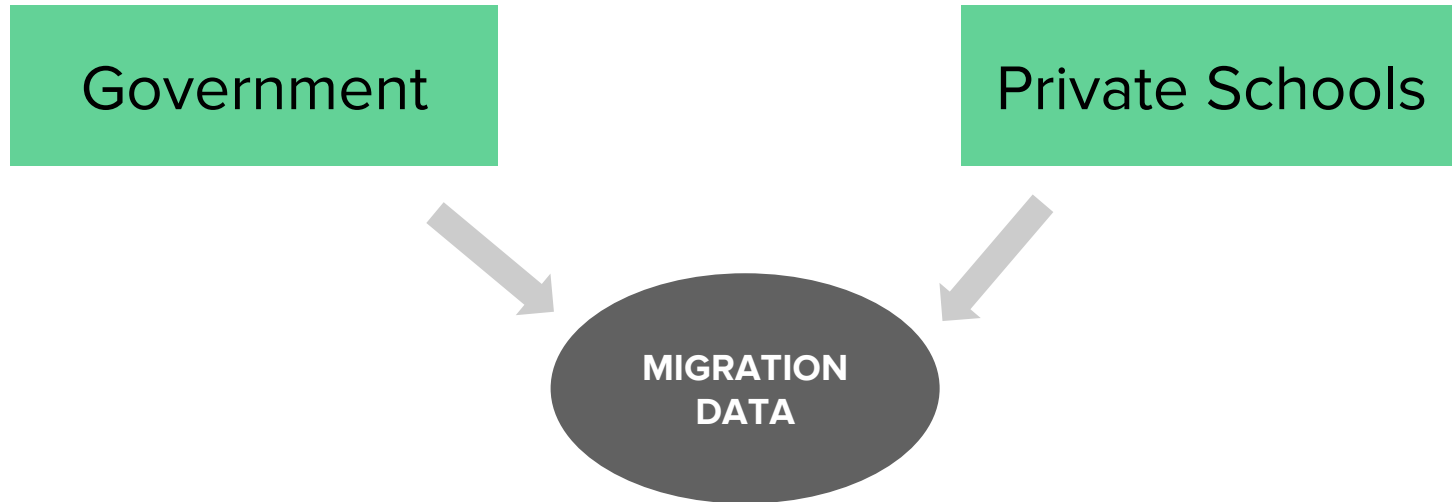
CPSC 547 - Final Presentation - December 2015

# Context

- Elementary and secondary education in Brazil consists of twelve years of education for children aged from 6 to 18 years (grades 1 to 12).
- There are both public and private schools available, the private ones usually outperforming public ones regarding education quality [1, 2].
- Public and private compete with each other; when quality of public education increases, there is a decrease in the number of enrollments in private schools [1].
- There is also a large number of private schools sharing the market and competing with each other for students to survive.

# Stakeholders

Understanding migration is useful for both government and private schools, because it may help them identify issues and potential areas of improvement.



# Tasks

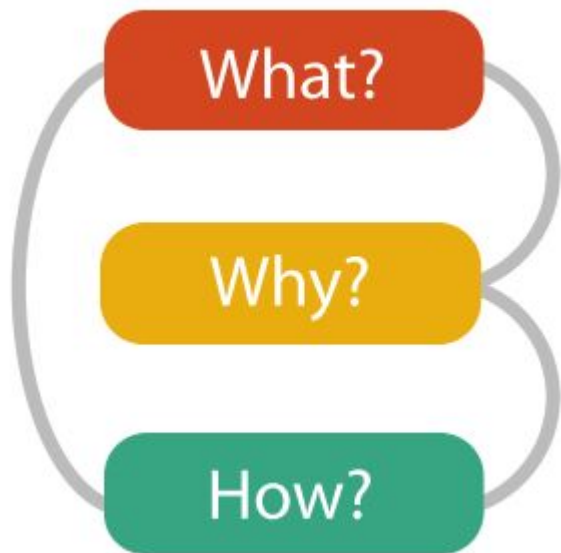
## Task for schools

- Is there any particular grade in which migration is more intense (losing/gaining more students)?
- To which schools are their students going to/coming from?
- Is it possible to identify a pattern in the geographic location of the main competitor schools?

## Task for government

- Are students migrating from public to private schools (or vice versa)?
- Which schools are gaining/losing more students per grade?
- Which grade is gaining/losing more students across schools?

# Data and Task Abstraction



## Dataset

- Educational Census, released every year and public available.
- For each student of the country: which school they were studying in that year, in which grade they were, and the school type (private or public), among other.
- Huge dataset; just for São Paulo state we have 10,581,500 students and 28,718 schools, as of 2014. The actual number is larger as we used data from census since 2012 until 2014, in the order of 10,000,000 rows each.

# What?

## Datasets

### → Data Types

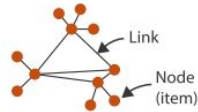
→ Items → Attributes → Links → Positions

### → Data and Dataset Types

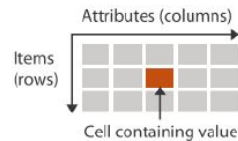
Networks & Trees	Tables	Geometry
Items (nodes)	Items	Items
Links	Attributes	Positions
Attributes		

### → Dataset Types

→ Networks



→ Tables



→ Geometry (Spatial)



### → Dataset Availability

→ Static



## Attributes

### → Attribute Types

→ Categorical



→ Ordered

→ Quantitative



### → Ordering Direction

→ Sequential



→ Diverging



# Why?

## Actions

## Targets

### → Analyze

→ Consume

→ Produce

→ *Discover*

→ *Derive*



### → Search

 *Locate*

### → Query

→ Identify

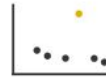
→ Compare

→ Summarize



### → All Data

→ Outliers



### → Network Data

→ Topology



# How?

## Encode

### → Arrange

→ Express



→ Order



→ Use



### → Map

from **categorical** and **ordered** attributes

→ Color

→ Hue



→ Size, Angle, Curvature, ...



→ Shape



## Manipulate

### → Select



### → Navigate

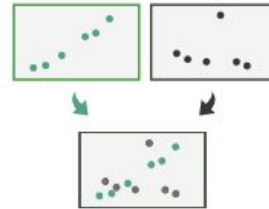


## Facet

### → Juxtapose



### → Superimpose



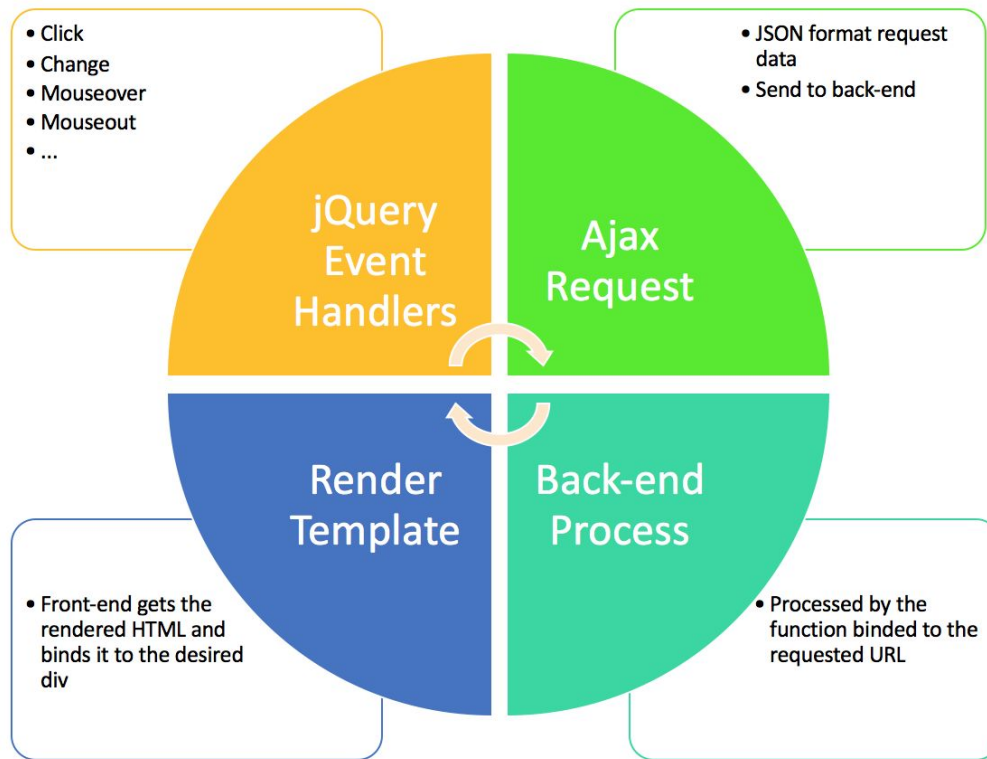
## Reduce

### → Filter





# Implementation



# Demo

<https://youtu.be/aYiPELe2sZs>

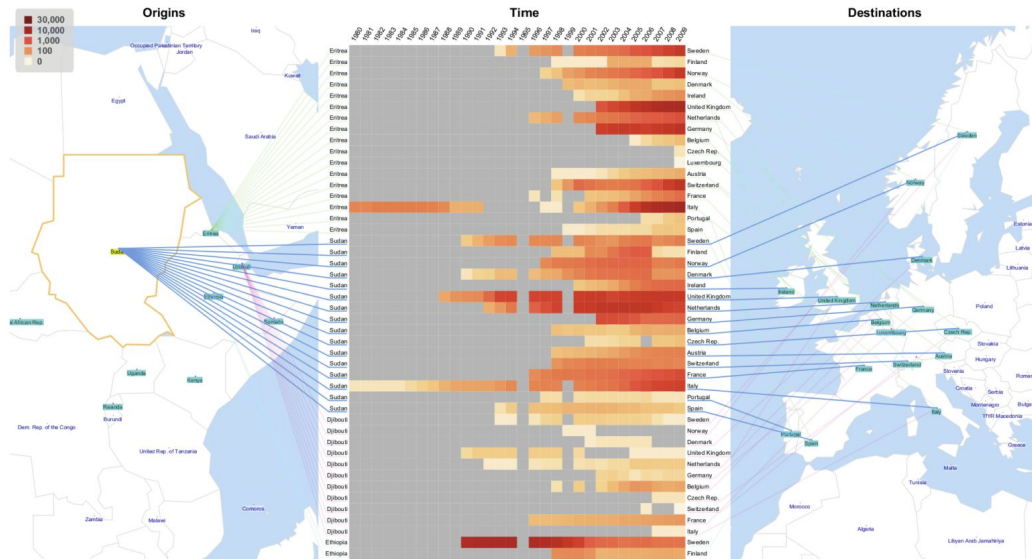
# Limitations and Critiques

## Overview panel

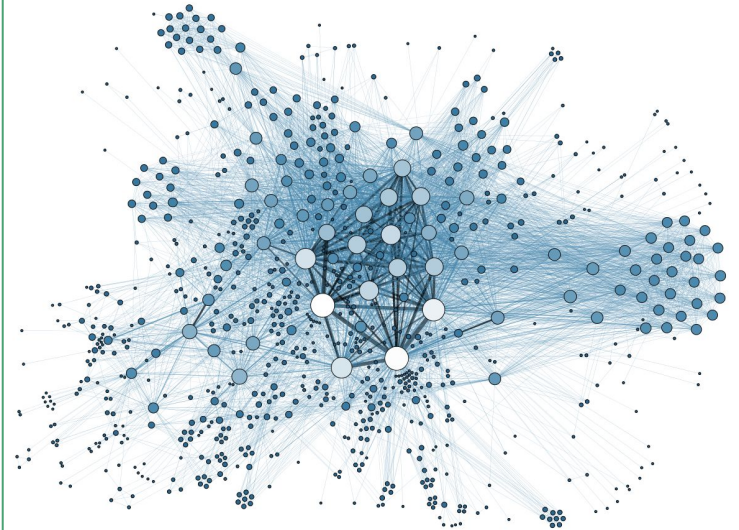
- Prioritized ranking and details per grades as tasks demanded. Trade-off years-grades, as we just had 3 years of data, we opted to not show progression among years.
- Interface familiar to users is probably going to be easy to understand to a non-computer savvy public.
- However, shows only small slices of data and do not offer a good way to compare migration among years or to identify patterns.
- More elaborated overview visualizations can be developed to explore further the potential of this dataset
  - regarding the identification of migration patterns among years
  - clustering schools into groups based on similar flow patterns

# Future Work

## Migration among years [3]



## Clusters of schools [4]



# Lessons learned

- Data cleaning and wrangling: time consuming, should plan for that.
- Processing times: can be very long when dealing with large data; try to test with samples first.
- Try to see beyond users demands and do not be attached to visualizations they are used to; prioritizing rankings and grades in the overview panel did not get all the potential this dataset had.
- Test early, before coding: Hard to predict all needed interactions in conceptual design and when drawing the mockups. Also hard to predict how well the visualization is going to meet users needs. Test before committing with coding effort.

# References

- [1] Estevan, Fernanda. “Public Education Expenditures and Private School Enrollment.” *Canadian Journal of Economics/Revue canadienne d'économique* (2015): Web.
- [2] Vandenberghe, V., and S. Robin. “Evaluating The Effectiveness of Private Education across Countries: a Comparison of Methods.” *Labour Economics* 11.4 (2004): 487–506. Web.
- [3] Boyandin, Ilya et al. “Flowstrates: An Approach For Visual Exploration of Temporal Origin-Destination Data.” *Computer Graphics Forum* 30.3 (2011): 971–980. Web.
- [4] By Martin Grandjean [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons. Available at [https://commons.wikimedia.org/wiki/File%3ASocial\\_Network\\_Analysis\\_Visualization.png](https://commons.wikimedia.org/wiki/File%3ASocial_Network_Analysis_Visualization.png)

Thank you!  
Any Questions?

Carolina: [carolamigo@gmail.com](mailto:carolamigo@gmail.com)

Dylan: [wdong@cs.ubc.ca](mailto:wdong@cs.ubc.ca)