

ĐỒ ÁN CUỐI KỲ TRỰC QUAN HOÁ DỮ LIỆU

THỐNG KÊ SỐ LIỆU GIÁO DỤC TIỂU HỌC

NHÓM 3

Nhóm 3

- 19120057 - Lê Quốc Cường
- 19120096 - Lưu Gia Minh
- 19120142 - Nguyễn Thị Phương Trang
- 19120145 - Lê Đào Duy Trọng
- 19120202 - Võ Tiến Dũng

Nội dung trình bày

01

Giới thiệu đề tài

02

Tìm hiểu dữ liệu

03

Tiền xử lý dữ liệu

04

Trực quan dữ liệu

05

Câu chuyện



01

Giới thiệu đề tài

Tổng quan về đề tài

Giáo dục tiểu học là gì?

Bậc học Tiểu học là bậc học đầu tiên trong hệ thống giáo dục phổ thông. Giáo dục tiểu học thực hiện trong 5 năm học, từ lớp 1 đến lớp 5. Đây là bậc học đóng vai trò nền tảng quan trọng đối với trẻ em.



Lý do chọn đề tài?

Giáo dục tiểu học nói riêng và giáo dục nói chung đã, đang và sẽ luôn là một trong những vấn đề được quan tâm hàng đầu.

Chọn tìm hiểu đề tài này nhằm biết thêm nhiều thông tin về sự phát triển giáo dục tiểu học trong thời gian qua.



02

Tìm hiểu dữ liệu

Nguồn dữ liệu

Tổng cục Thống kê Việt Nam: <https://www.gso.gov.vn/>

Các tập dữ liệu lấy bao gồm:

Số trường phổ thông tại thời điểm 30/9 phân theo địa phương

Số lớp học phổ thông tại thời điểm 30/9 phân theo địa phương

Số học sinh phổ thông tại thời điểm 30/9 phân theo địa phương

Số học sinh phổ thông thuộc các dân tộc thiểu số tại thời điểm 30/9 phân theo địa phương

Số nữ giáo viên phổ thông trực tiếp giảng dạy tại thời điểm 30/9 phân theo địa phương

Thu thập dữ liệu

Lấy data của các trường dữ liệu qua các năm 2017 - 2020 trên <https://www.gso.gov.vn> lưu thành 1 file excel gồm nhiều sheet tương ứng với các trường dữ liệu.

Số lượng lớp học

	Địa phương	2017	2018	2019	2020
1	CẢ NƯỚC	279974	280179	281565	284084
2	Đồng bằng sông Hồng	55437	57036	58357	60123
3	Hà Nội	17116	18238	19048	19958
4	Hà Tây	NaN	NaN	NaN	NaN
5	Vĩnh Phúc	3237	3420	3572	3687

Số lượng học sinh

	Địa phương	2017	2018	2019	2020
1	CẢ NƯỚC	8041842	8541451	8741545	8891344
2	Đồng bằng sông Hồng	1860472	2021081	2083595	2155722
3	Hà Nội	678739	738864	765571	791547
4	Hà Tây	NaN	NaN	NaN	NaN
5	Vĩnh Phúc	105801	116240	121686	125578

Cấu trúc bảng dữ liệu

	Địa phương	Năm	Số trường học	Số lớp học	Số học sinh	Số học sinh thiếu số	Số giáo viên nữ và học sinh nữ
1	Hà Nội	2017	727	17116	678739	7812	23228
2	Hà Tây	2017	0	0	0	0	0
3	Vĩnh Phúc	2017	176	3237	105801	5955	4046
4	Bắc Ninh	2017	155	3203	106948	0	4363
5	Quảng Ninh	2017	180	4275	116355	15094	5275

Chi tiết về dữ liệu

Tên thuộc tính	Kiểu dữ liệu	Ý nghĩa
Địa phương	string	Tên tỉnh thành/ vùng
Năm	int	Năm thống kê
Số trường học	int	Số lượng trường học
Số lớp học	int	Số lượng lớp học
Số học sinh	int	Số lượng học sinh
Số học sinh thiếu số	int	Số lượng học sinh thiếu số
Số giáo viên nữ	int	Số lượng giáo viên nữ



03

Tiền xử lý dữ liệu

Trích xuất dữ liệu

Từ các bộ giá trị theo năm tìm được, ta thực hiện lọc dữ liệu ở các dataframe. Thêm vào cột **Năm**, và thứ tự các cột dữ liệu còn lại lần lượt là: **Số trường học, Số lớp học, Số học sinh, Số học sinh thiếu số, Số giáo viên nữ.**

Xử lý dữ liệu

Tạo các bộ dữ liệu theo từng năm từ việc trích xuất các bộ dữ liệu vừa thu thập

Ghép các bộ dữ liệu mới tạo được thành 1 data hoàn chỉnh. Tuy có sự trùng lặp ở cột Địa phương và Năm nhưng sẽ dễ hơn khi thao tác với công cụ trực quan.

	Địa phương	Năm	Số trường học	Số lớp học	Số học sinh	Số học sinh thiếu số	Số giáo viên nữ và học sinh nữ
0	CẢ NƯỚC	2017	14937	279974	8041842	1416710	309815
1	Đồng bằng sông Hồng	2017	2770	55437	1860472	NaN	72041
2	Hà Nội	2017	727	17116	678739	7812	23228
3	Hà Tây	2017	NaN	NaN	NaN	NaN	NaN
4	Vĩnh Phúc	2017	176	3237	105801	5955	4046
...
279	Cần Thơ	2020	173	3114	98079	2968	2930
280	Hậu Giang	2020	153	2406	67192	2778	1613
281	Sóc Trăng	2020	209	4064	119456	49559	2603
282	Bạc Liêu	2020	118	2336	74721	8914	1750
283	Cà Mau	2020	222	3823	109690	4175	3332

Bộ dữ liệu

Sheet **total_df** có 252 dòng với số liệu của 63 tỉnh thành trong 4 năm 2017 - 2020.

Sheet **area_df** có 28 dòng với số liệu của 7 khu vực (tính luôn Cả nước) trong 4 năm 2017 - 2020.

	Địa phương	Năm	Số trường học	Số lớp học	Số học sinh	Số học sinh thiếu số	Số giáo viên nữ
1	Hà Nội	2017	727	17116	678739	7812	23228
2	Hà Tây	2017	0	0	0	0	0
3	Vĩnh Phúc	2017	176	3237	105801	5955	4046
4	Bắc Ninh	2017	155	3203	106948	0	4363
5	Quảng Ninh	2017	180	4275	116355	15094	5275

	Địa phương	Năm	Số trường học	Số lớp học	Số học sinh	Số học sinh thiếu số	Số giáo viên nữ
0	CẢ NƯỚC	2017	14937.0	279974	8041842	1416710.0	309815.0
1	Đồng bằng sông Hồng	2017	2770.0	55437	1860472	NaN	72041.0
2	Trung du và miền núi phía Bắc	2017	2681.0	51036	1160871	NaN	54198.0
3	Bắc Trung Bộ và Duyên hải miền Trung	2017	3675.0	60357	1622187	NaN	70232.0
4	Tây Nguyên	2017	1252.0	22512	592671	NaN	26160.0
5	Đông Nam Bộ	2017	1522.0	37576	1334462	NaN	42751.0
6	Đồng bằng sông Cửu Long	2017	3037.0	53056	1471179	NaN	44433.0

Thống kê trung bình của các trường dữ liệu qua từng năm.

<div>Trường dữ liệu</div> <div>Năm</div>	2017	2018	2019	2020
Số trường học	233		207	201
Số lớp học	4374	4447	4469	4509
Số học sinh	125653	148727	138754	14132
Số học sinh thiếu số	22105	23915	24565	25346
Số giáo viên nữ	4840		4756	4813



04

Trực quan dữ liệu



Công cụ trực quan hóa



Tableau

Tableau là phần mềm do công ty Tableau Software xây dựng và phát triển. Phần mềm hỗ trợ phân tích và trực quan hoá dữ liệu.

Tableau có thể giúp người dùng xây dựng các **biểu đồ**, **đồ thị**, **bản đồ**, xây dựng **dashboard** cũng các **câu chuyện** từ dữ liệu.



Seaborn

Seaborn là một thư viện trực quan hóa dữ liệu của Python dựa trên **Matplotlib** cung cấp nhiều loại biểu đồ đa dạng để vẽ và mô phỏng dữ liệu.





Các loại biểu đồ sử dụng

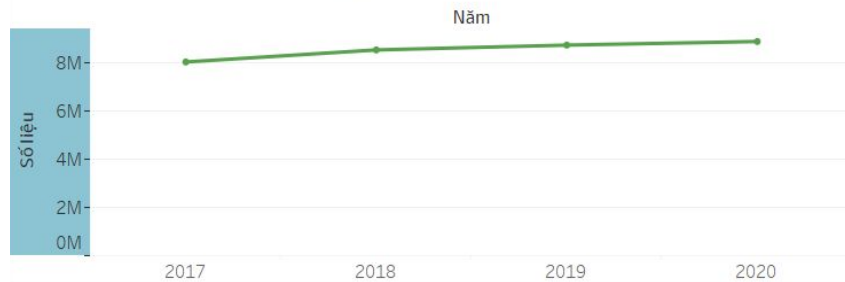


Biểu đồ

Số liệu cả nước theo thời gian:

- **Loại biểu đồ:** Biểu đồ đường
- **Trục X:** Năm
- **Trục Y:** Số liệu tương ứng trường dữ liệu
- **Khả năng tương tác:**
 - Trường dữ liệu

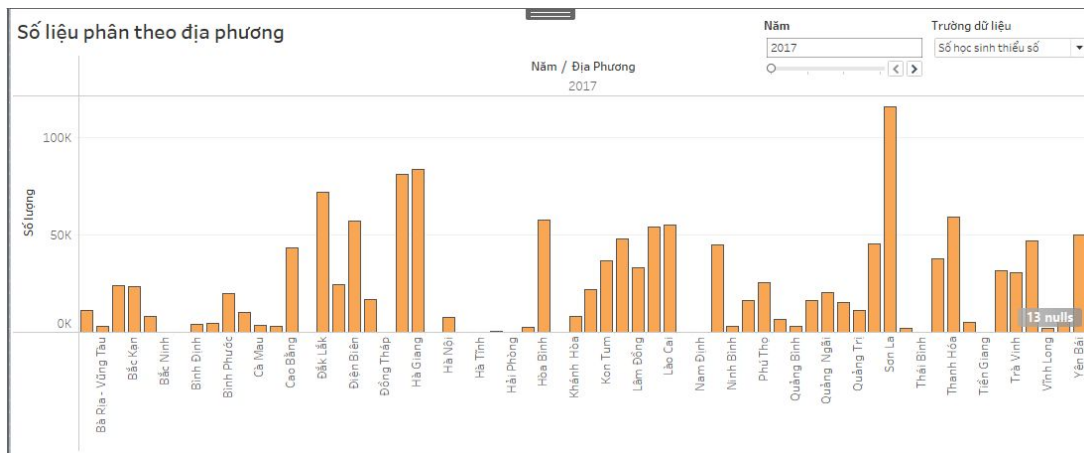
Số liệu cả nước theo thời gian



Biểu đồ

Số liệu phân theo địa phương

- **Loại biểu đồ:** Biểu đồ cột
- **Trục X:** Các tỉnh/thành phố
- **Trục Y:** Số liệu thu thập ở mỗi địa phương tương ứng
- **Khả năng tương tác:**
 - Trường dữ liệu
 - Năm

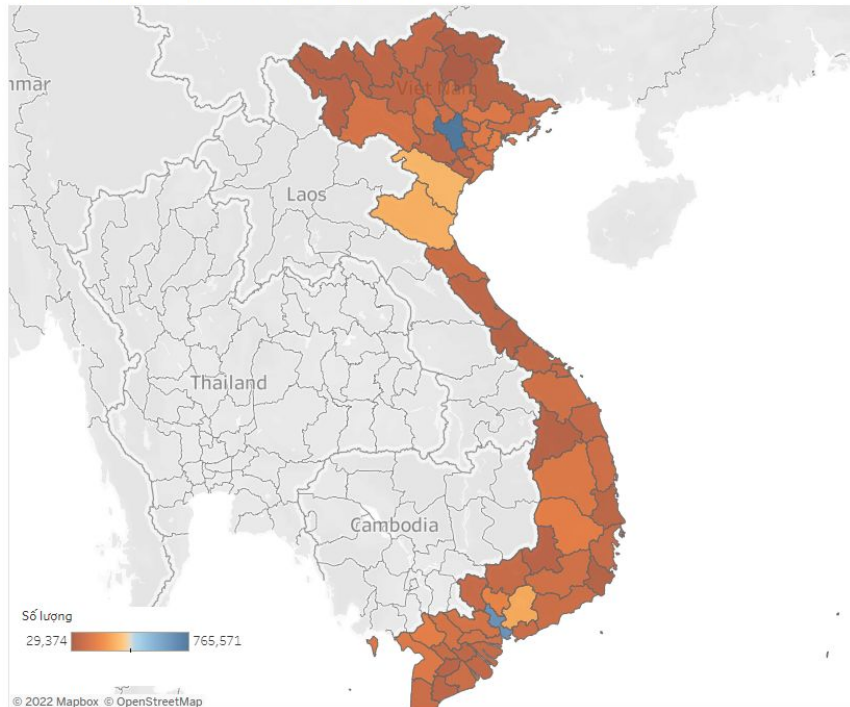


Biểu đồ

Biểu đồ mật độ phân bố của trường dữ liệu

- **Loại biểu đồ:** Heatmap theo bản đồ VN
- **Màu sắc:** Thể hiện cho số liệu theo trường thông tin được chọn theo năm
- **Khả năng tương tác:**
 - Trường dữ liệu
 - Năm

Mật độ phân bố của Số học sinh năm 2019

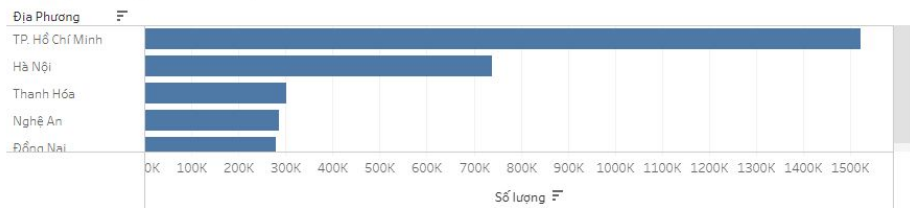


Biểu đồ

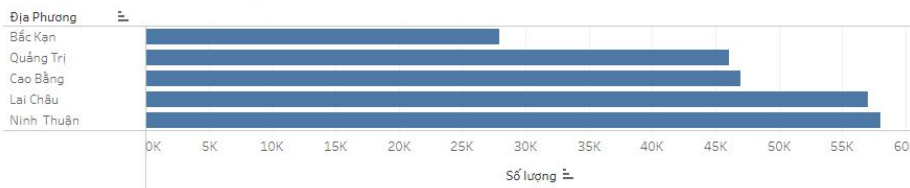
Biểu đồ N tỉnh/ thành phố có giá trị lớn nhất/ nhỏ nhất theo năm

- **Loại biểu đồ:** Biểu đồ cột
- **Trục X:** Số liệu tương ứng trường dữ liệu
- **Trục Y:** Địa phương
- **Khả năng tương tác:**
 - Trường dữ liệu
 - Năm
 - Giá trị top muốn lấy

5 tỉnh/thành phố có Số học sinh lớn nhất năm 2018

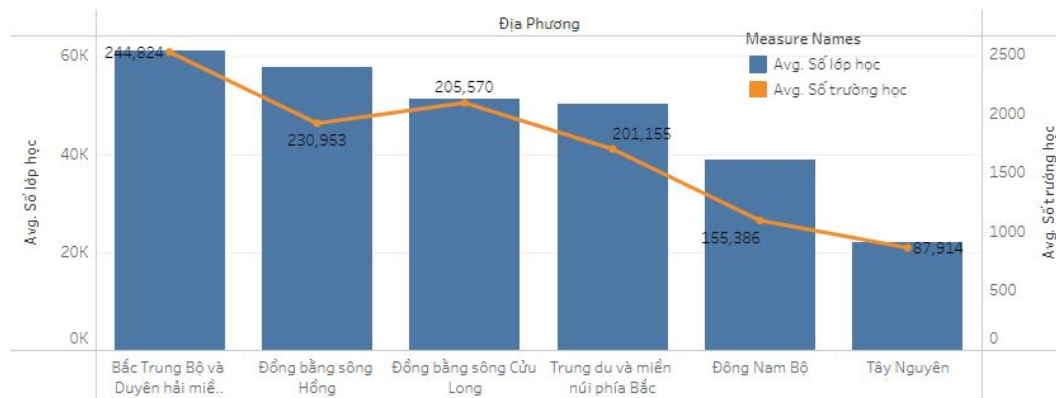


5 tỉnh/thành phố có Số học sinh nhỏ nhất năm 2018



Biểu đồ

Phân bố số lớp học và số trường học theo vùng

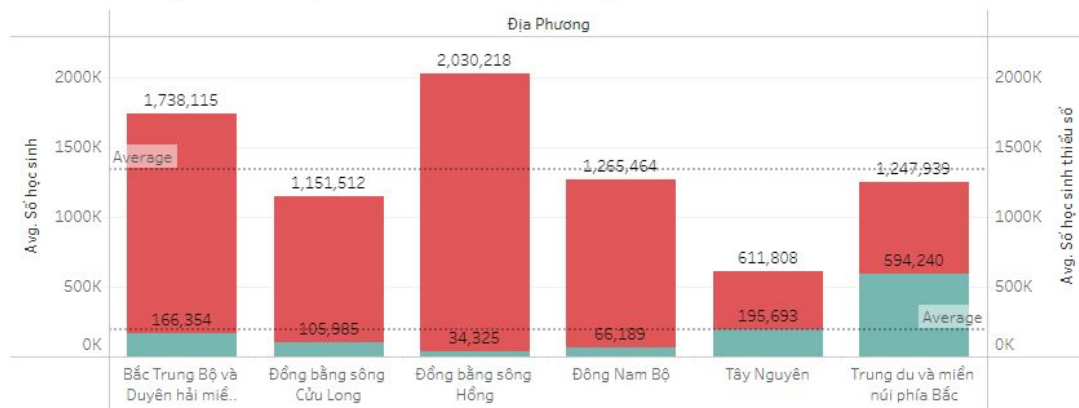


Biểu đồ cột (trực quan số lớp học) kết hợp biểu đồ đường (trực quan số trường học)

- **Loại biểu đồ:** Biểu đồ tích hợp (cột và đường) (duo axis)
- **Trục X:** Vùng địa phương
- **Trục Y:** Trung bình số lớp học (trái), Trung bình số trường học (phải)
- **Khả năng tương tác:**
 - Trường dữ liệu

Biểu đồ

Phân bố số học sinh và học sinh thiếu số theo vùng



Biểu đồ cột thể hiện số học sinh và số học sinh thiếu số theo vùng

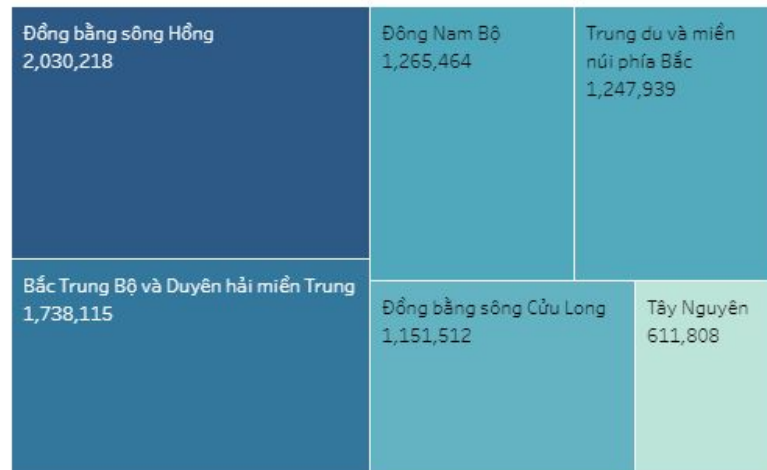
- **Loại biểu đồ:** Biểu đồ cột đôi (duo axis + synchronise axis)
- **Trục X:** Vùng địa phương
- **Trục Y:**
 - Trung bình 4 năm của số học sinh (trái)
 - Trung bình 4 năm của số học sinh thiếu số (phải)
- **Khả năng tương tác:**
 - Trường dữ liệu

Biểu đồ

Avg. Số học sinh



Số học sinh theo vùng



Biểu đồ treemap thể hiện số học sinh theo vùng

- **Loại biểu đồ:** Biểu đồ treemap
- **Trường dữ liệu:**
 - Vùng địa phương
 - Trung bình 4 năm của số học sinh
- **Khả năng tương tác:**
 - Trường dữ liệu

Biểu đồ 4

Số giáo viên nữ theo vùng

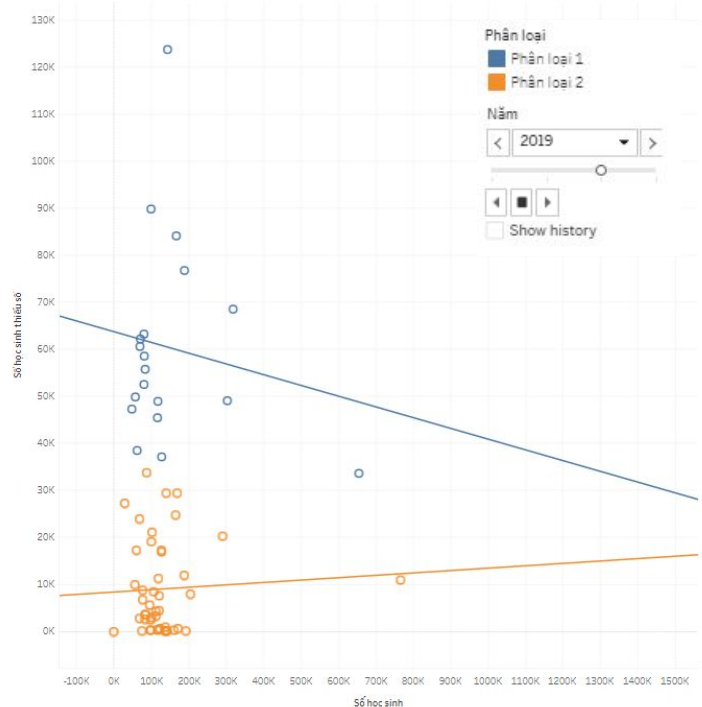


Biểu đồ packed bubbles thể hiện số học sinh và giáo viên nữ theo vùng

- **Loại biểu đồ:** Biểu đồ packed bubbles
- **Trường dữ liệu:**
 - Vùng địa phương
 - Trung bình 4 năm của số học sinh và giáo viên nữ
- **Khả năng tương tác:**
 - Trường dữ liệu

Biểu đồ

Phân phối các địa phương theo số học sinh và số học sinh thiếu số - 2019

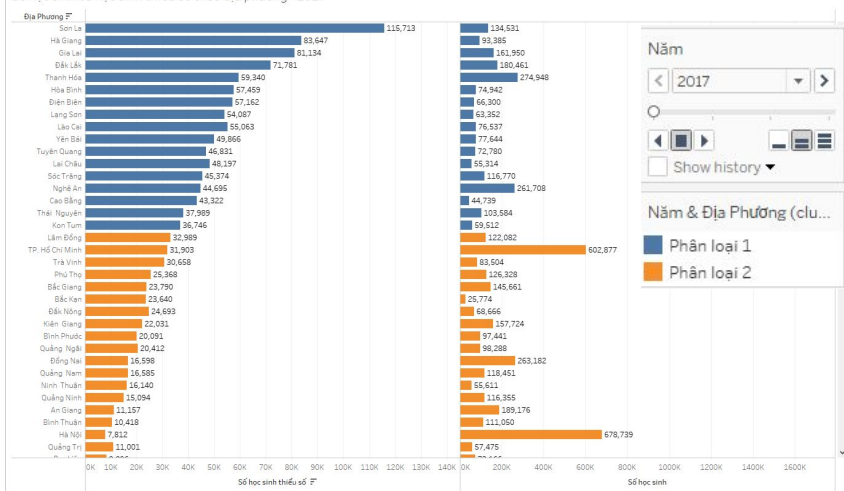


Biểu đồ scatter plot thể hiện sự phân bố của số học sinh và số học sinh thiếu số theo địa phương (Phân nhóm dựa trên mật độ phân bố dữ liệu)

- **Loại biểu đồ:** Biểu đồ scatter plot
- **Trục X:** Số học sinh
- **Trục Y:** Số học sinh thiếu số
- **Khả năng tương tác:**
 - Trường dữ liệu
 - Năm

Biểu đồ

Số học sinh và học sinh thiếu số theo địa phương - 2017



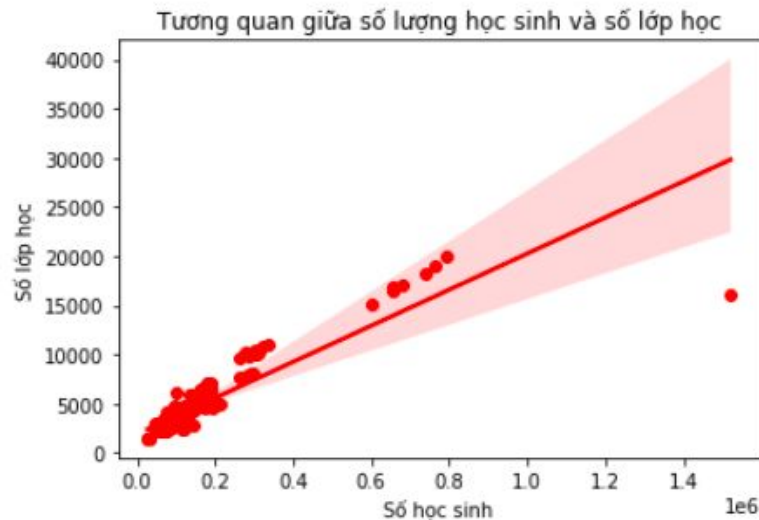
Biểu đồ cột đôi thể hiện sự phân bố của số học sinh và số học sinh thiếu số theo địa phương (Phân nhóm dựa trên mật độ phân bố dữ liệu)

- **Loại biểu đồ:** Biểu đồ cột đôi
- **Trục X1:** Số học sinh
- **Trục X2:** Số học sinh thiếu số
- **Trục Y:** địa phương
- **Khả năng tương tác:**
 - Trường dữ liệu
 - Năm

Biểu đồ

Biểu đồ tương quan giữa số lượng học sinh và số lượng lớp học

- **Loại biểu đồ:** Biểu đồ đường hồi quy
- **Trục X:** Số học sinh
- **Trục Y:** Số lớp học



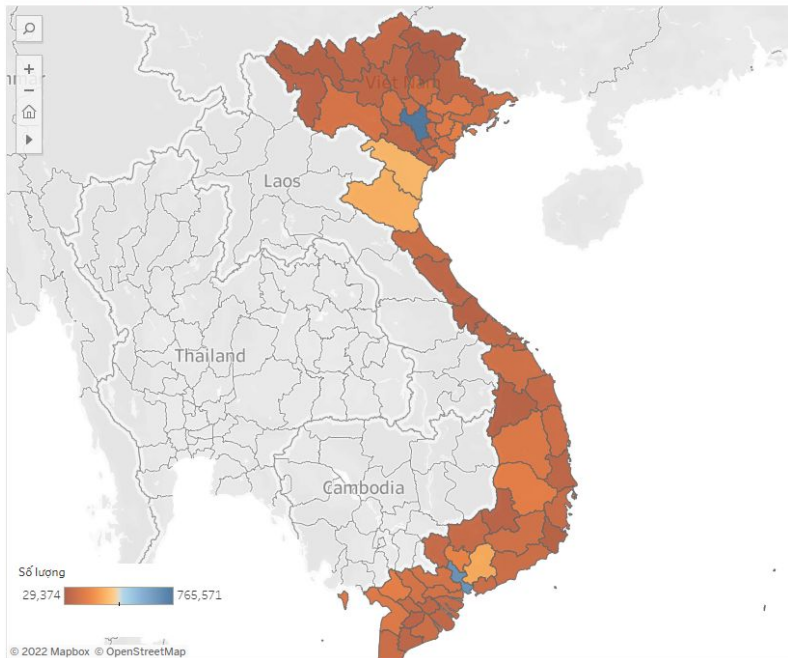


Các dashboard xây dựng

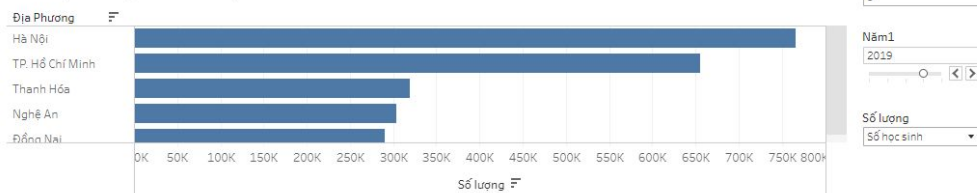


Dashboard Cả nước

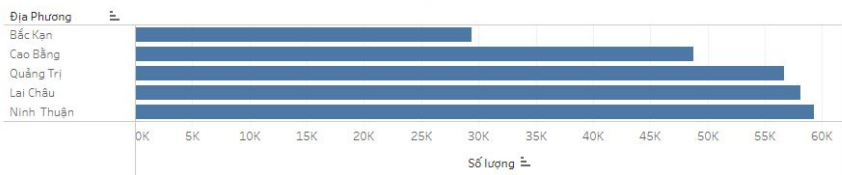
Mật độ phân bố của Số học sinh năm 2019



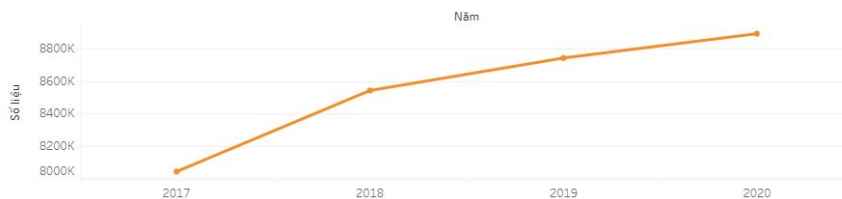
5 tỉnh/thành phố có Số học sinh lớn nhất năm 2019



5 tỉnh/thành phố có Số học sinh nhỏ nhất năm 2019



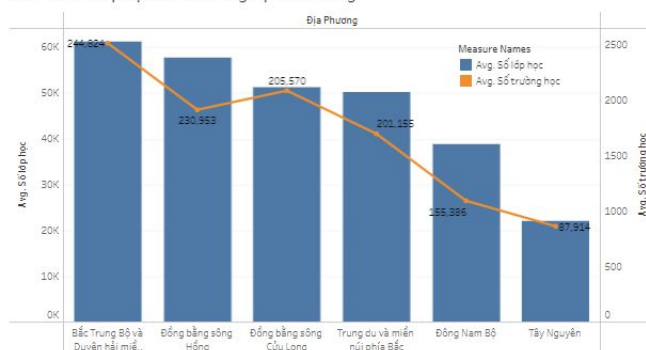
Số liệu Số học sinh cả nước theo thời gian



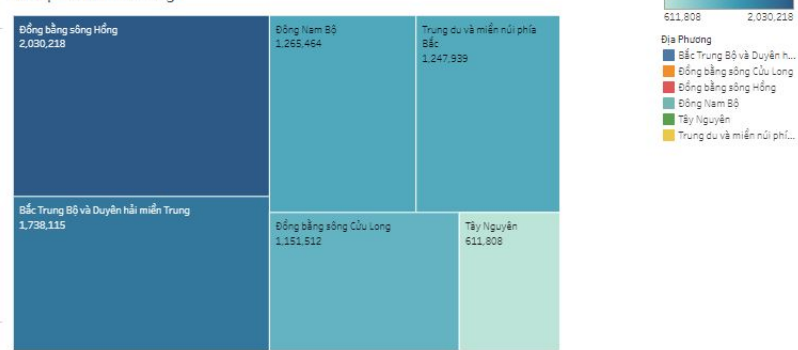
Dashboard thống kê số liệu giáo dục tiểu học theo vùng

DB_CN | DB_Vùng | DB_ | DB_Cluster

Phân bố số lớp học và số trường học theo vùng



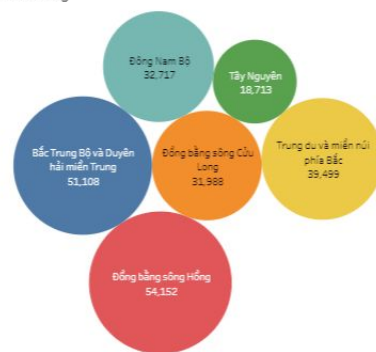
Số học sinh theo vùng



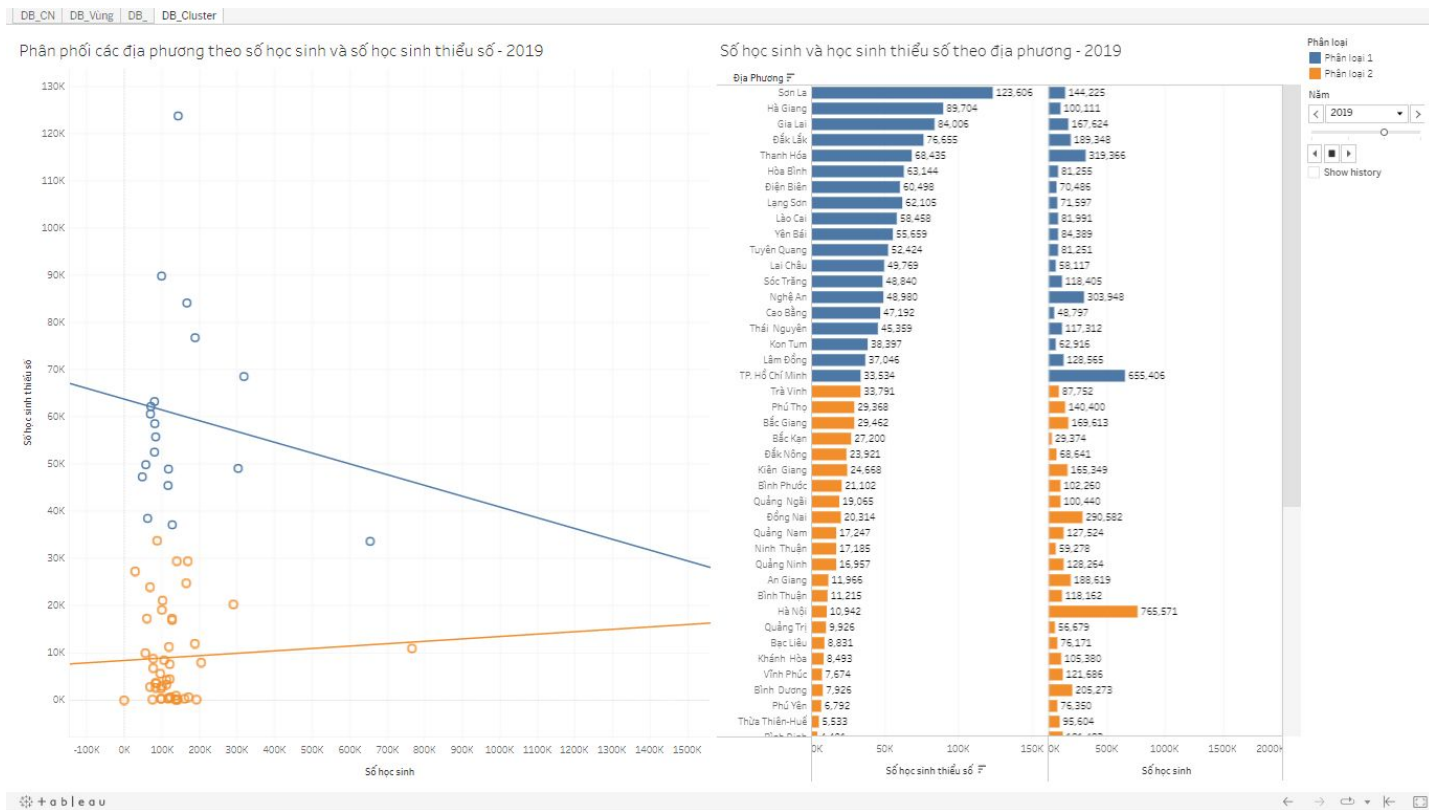
Phân bố số học sinh và học sinh thiếu số theo vùng



Số giáo viên nữ theo vùng



Dashboard phân nhóm các địa phương dựa trên sự phân bố của số học sinh thiếu số và số học sinh





Màu sắc



Đỏ, Cam và Xanh là 3 màu chủ đạo trong biểu đồ










Một gam màu nóng, làm nổi bật các thông tin đặc trưng của bộ dữ liệu, tạo sự thu hút cho người xem.



Vẫn là một gam màu nóng, làm nổi bật các thông tin nhưng còn mang lại cảm giác dịu nhẹ, cuốn hút người xem.



Một gam màu lạnh, tương phản với 2 màu còn lại, hỗ trợ làm nổi bật cho nhau và giúp cho người xem có thể phân biệt rõ ràng thông tin của dữ liệu.

	Đỏ (R)	Xanh lá (G)	Xanh dương (B)
Người thường			
Weak (Yếu)			
Blind (Mù)			

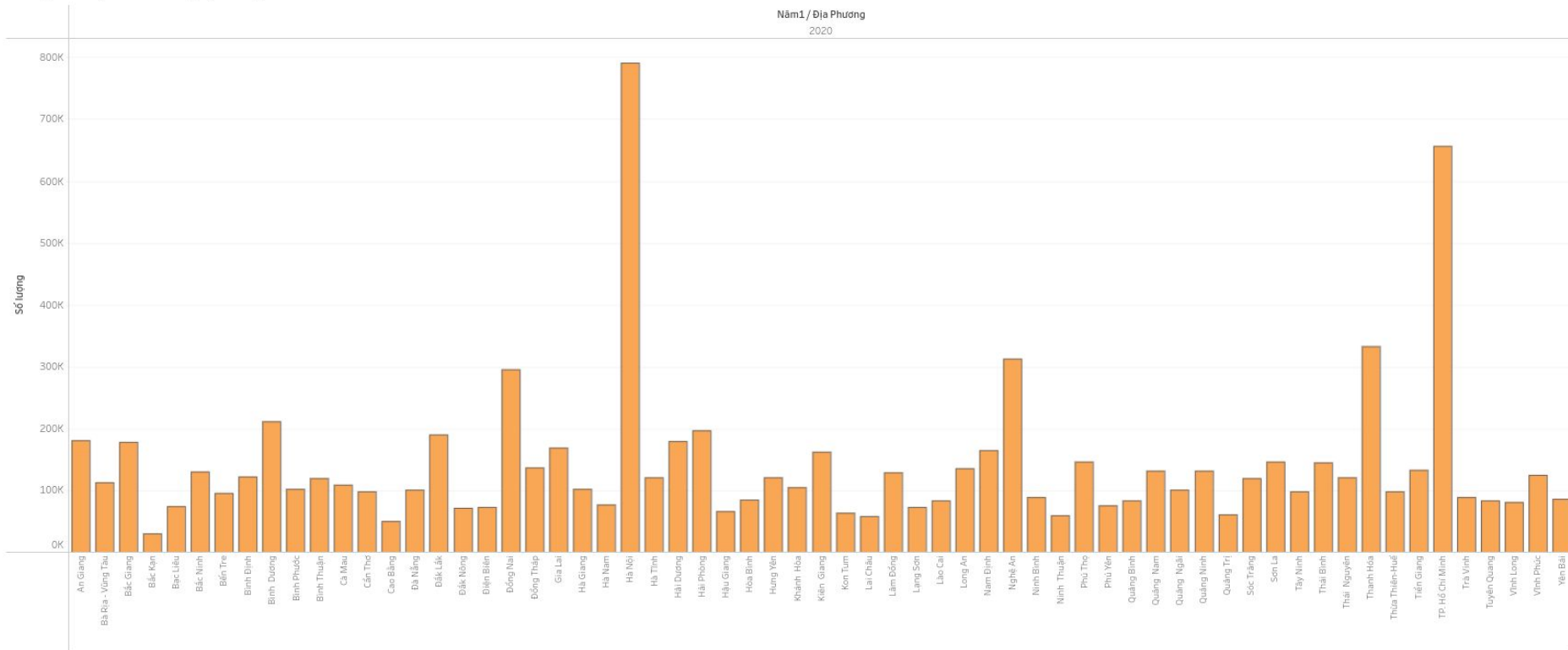


05

Câu chuyện

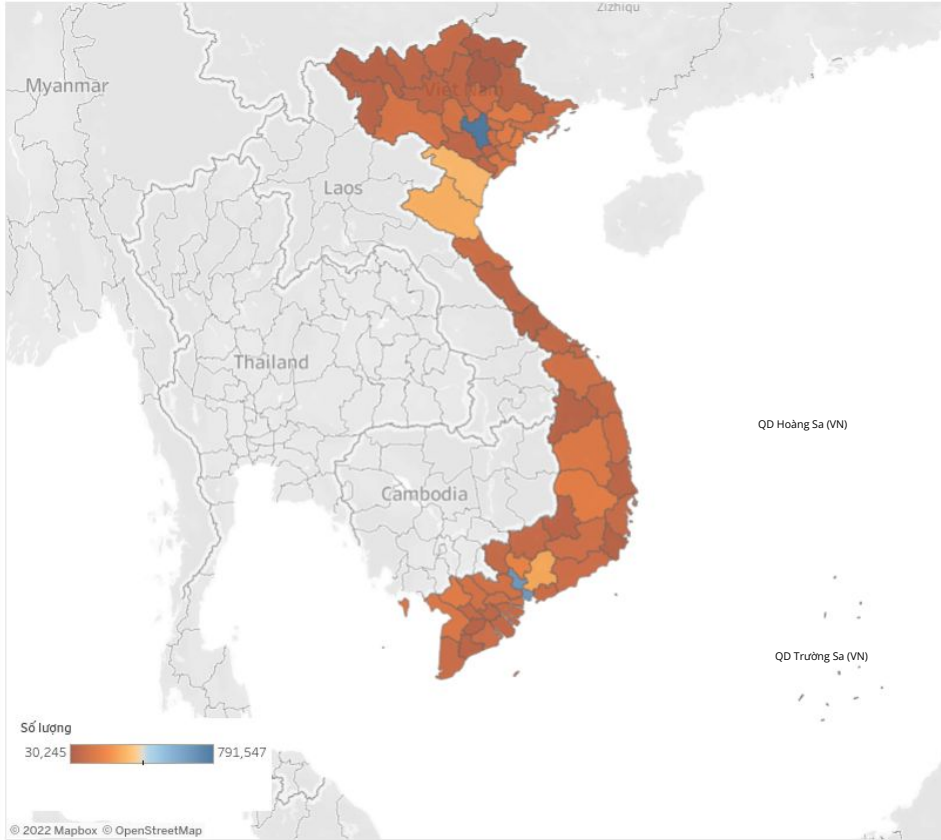
Biểu đồ

Số học sinh phân theo địa phương



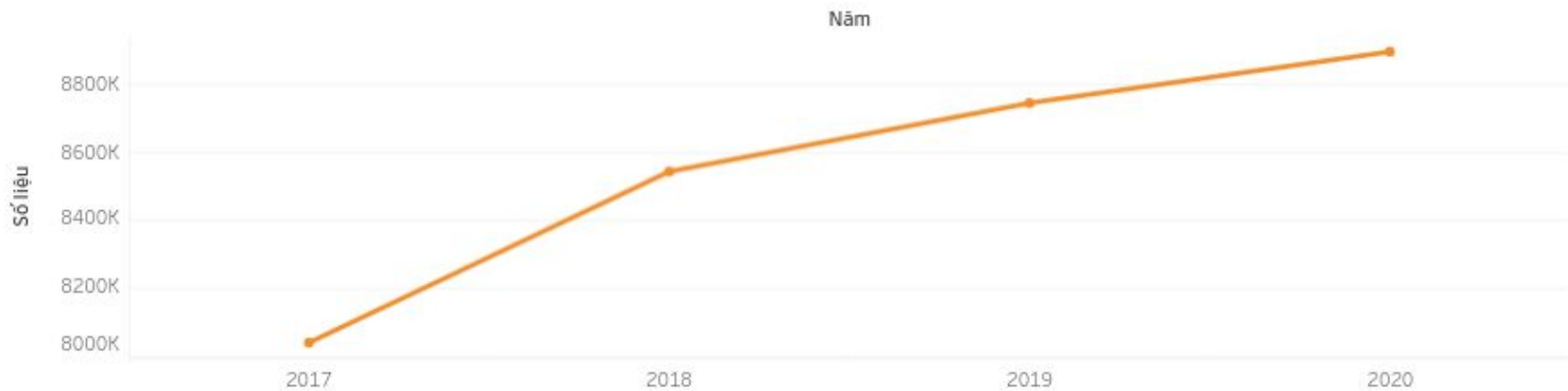
Biểu đồ

Mật độ phân bố của Số học sinh năm 2020

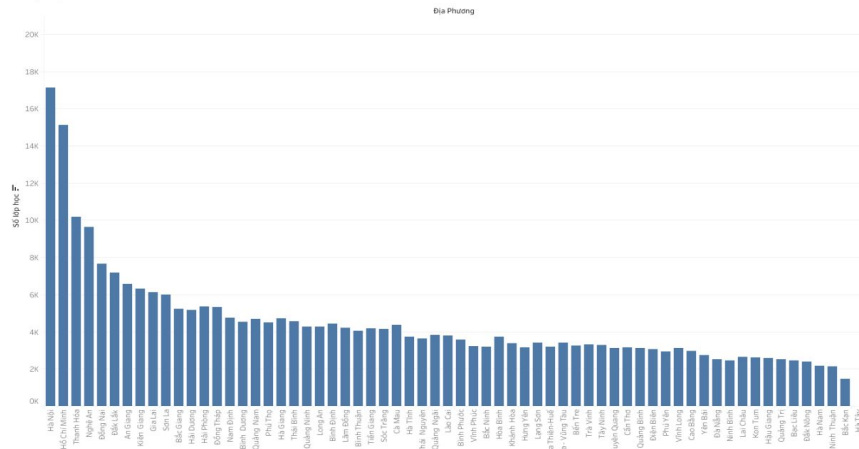


Biểu đồ

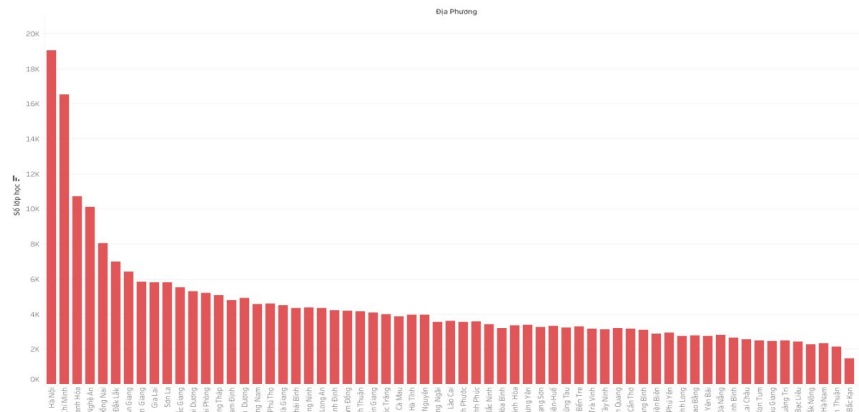
Số liệu Số học sinh cả nước theo thời gian



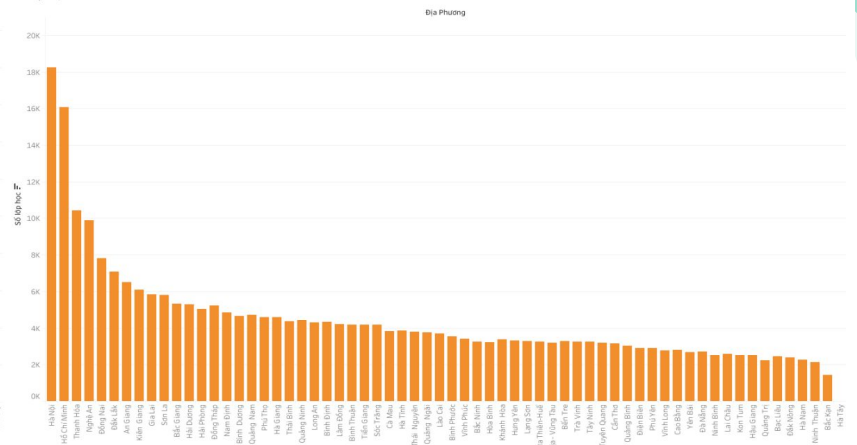
Số lớp học - 2017



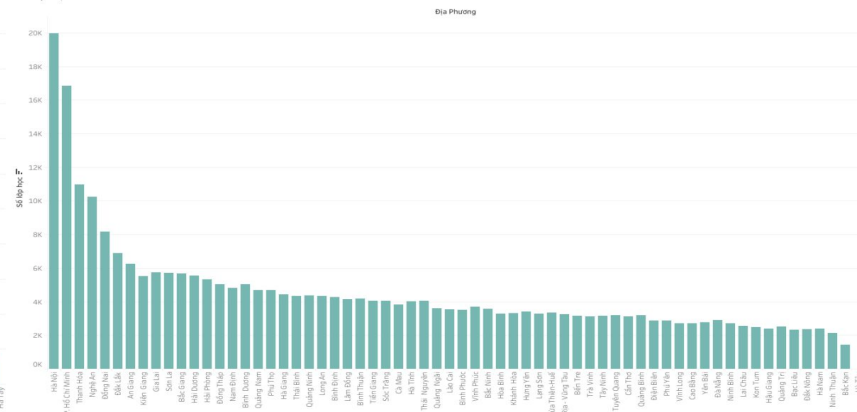
Số lớp học - 2019



Số lớp học - 2018

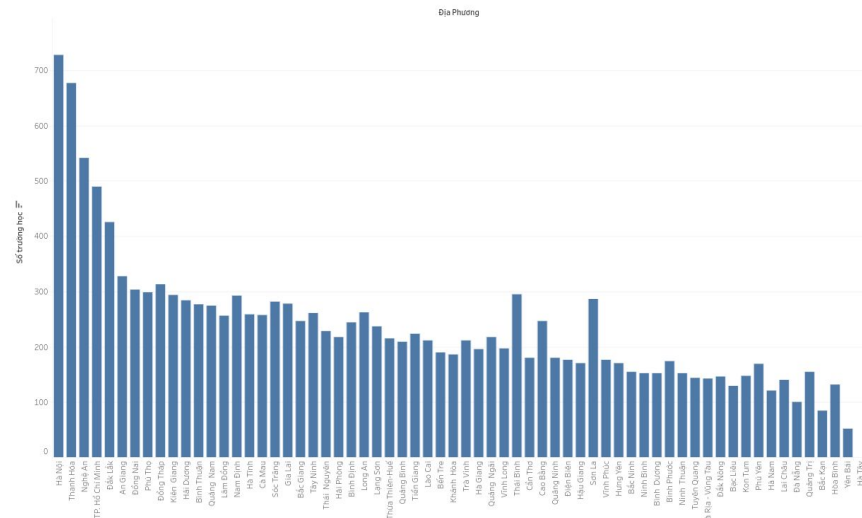


Số lớp học - 2020

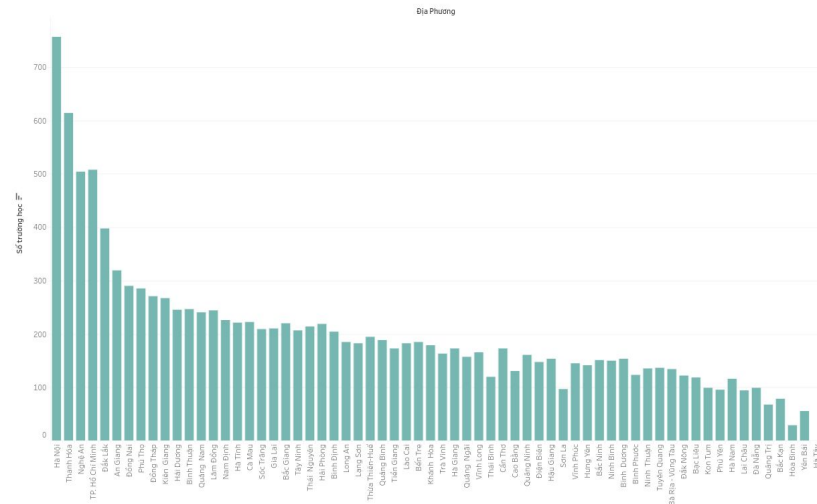


Biểu đồ

Số trường học - 2017

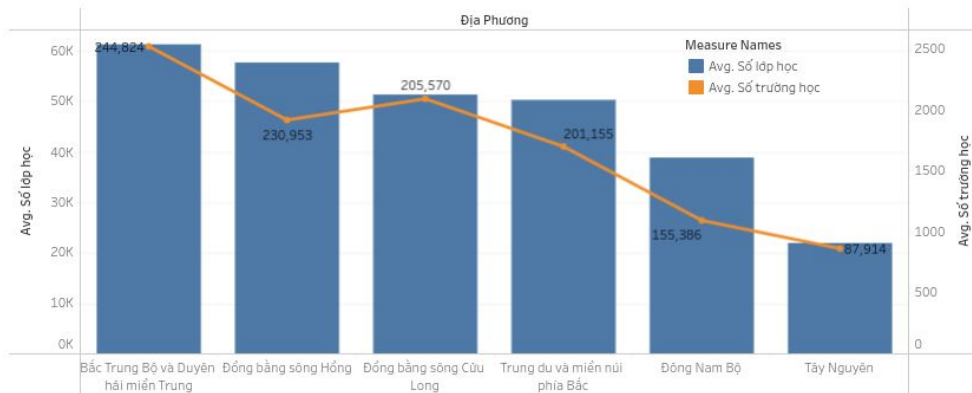


Số trường học - 2020



Biểu đồ

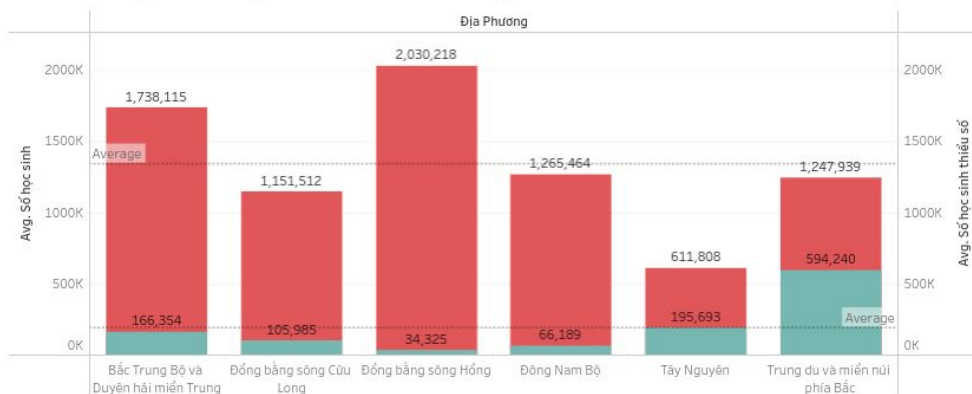
Phân bố số lớp học và số trường học theo vùng



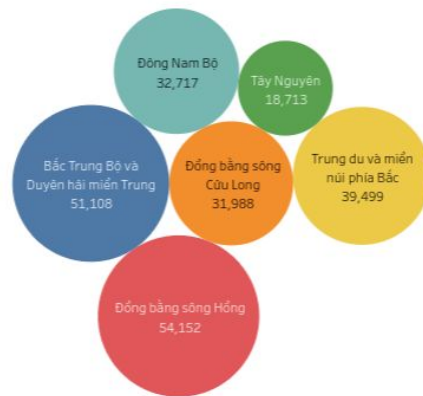
Số học sinh theo vùng



Phân bố số học sinh và học sinh thiếu số theo vùng



Số giáo viên nữ theo vùng





**Cảm ơn thầy
đã lắng nghe**