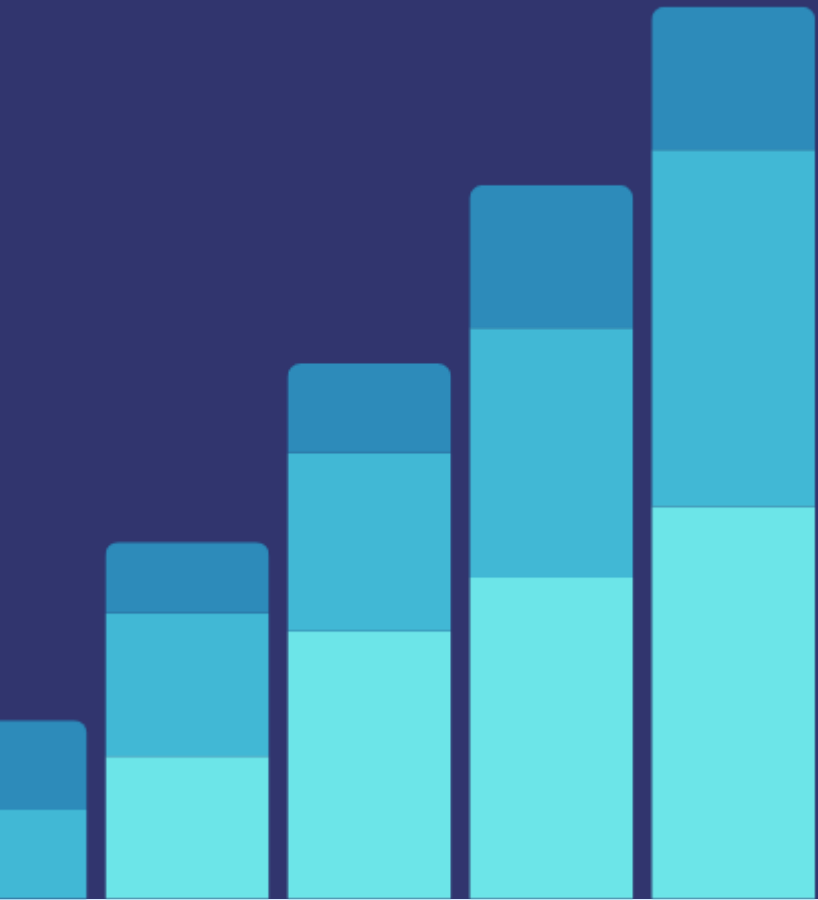


# **Data Engineering 1 – Term Project 2**

**Healthcare, Education, and GDP: Factors Behind Life Expectancy**

Elene Zuroshvili, Hamza Dhich, Nurbek Bektursyn, Matyas Kovacs

# Objective of the project



Our goal is to analyze the relationship between life expectancy and key socioeconomic factors, such as GDP, healthcare expenditure, and education spending, and to uncover insights into global development trends.

# Data sources for 2021

**RDBMS:  
World Database**

Source:  
MySQL Sample  
database

**CSV:  
Education  
Expenditure**

Source:  
Our world in data

**API:  
Health  
Expenditure**

Source:  
World Bank Group

**API:  
GDP**

Source:  
World Bank Group

A series of ten horizontal bars of varying lengths in a light blue color, positioned on the left side of the slide. The bars are arranged in a staggered, descending pattern from top to bottom, with the top bar being the longest and the bottom bar being the shortest.

**Let's explore our Knime  
workflow!**

# Loading the data

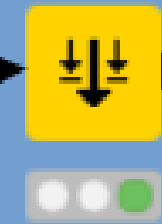
- Import the database to MySql workbench
- Connect MySql to Knime
- Query for the Countries table

- Download and read CSV
- Filter down to 2021

## 1. Load Countries from World database

MySQL Connector DB Query Reader

Column Filter



Connect to DB   Fetch All Countries   Remove Unneeded Columns

## 2. Loading file from file system

CSV Reader

Row Filter



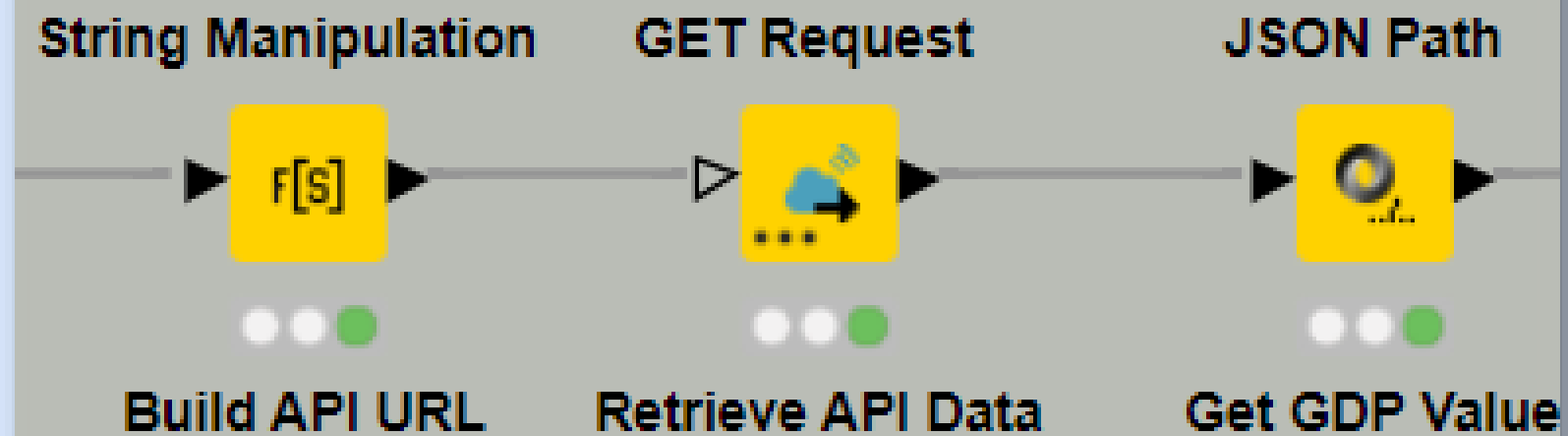
Load CSV File

Select Only 2021

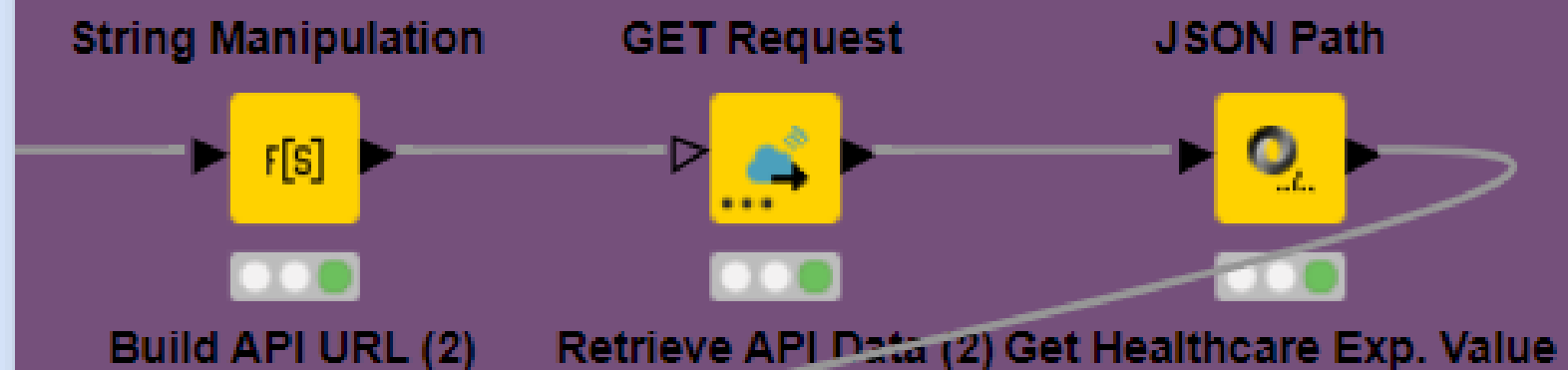
# API as a data source

Filter out the API's to all countries and year 2021 using Postman

## 3. Importing data from API1



## 4. Importing data from API2



Importing API's using Get Request and JSON Reader Nodes in Knime

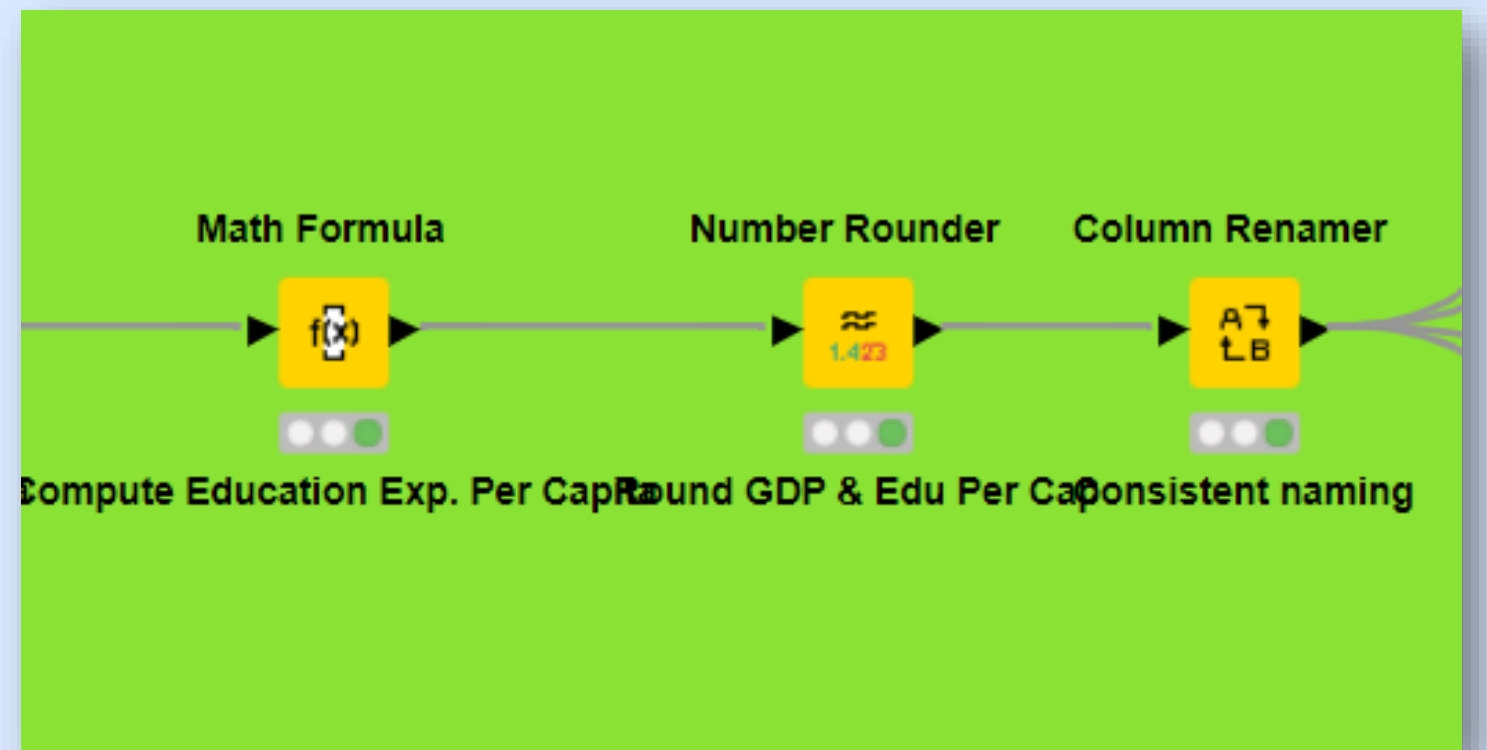
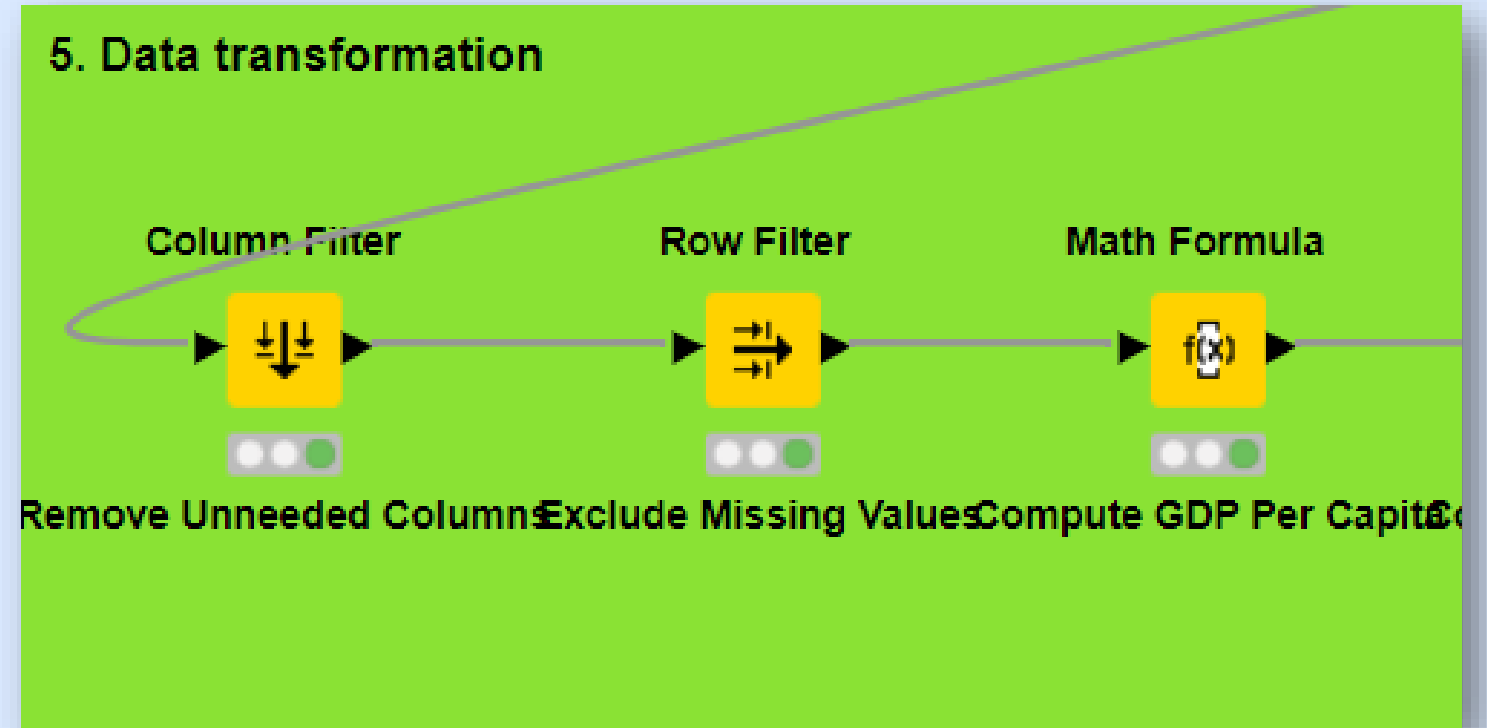
# Data transformation

Data preprocessing:

- Filtering columns
- Handling missing values

Feature Engineering:

- GDP per Capita
- Education Exp. Per Capita

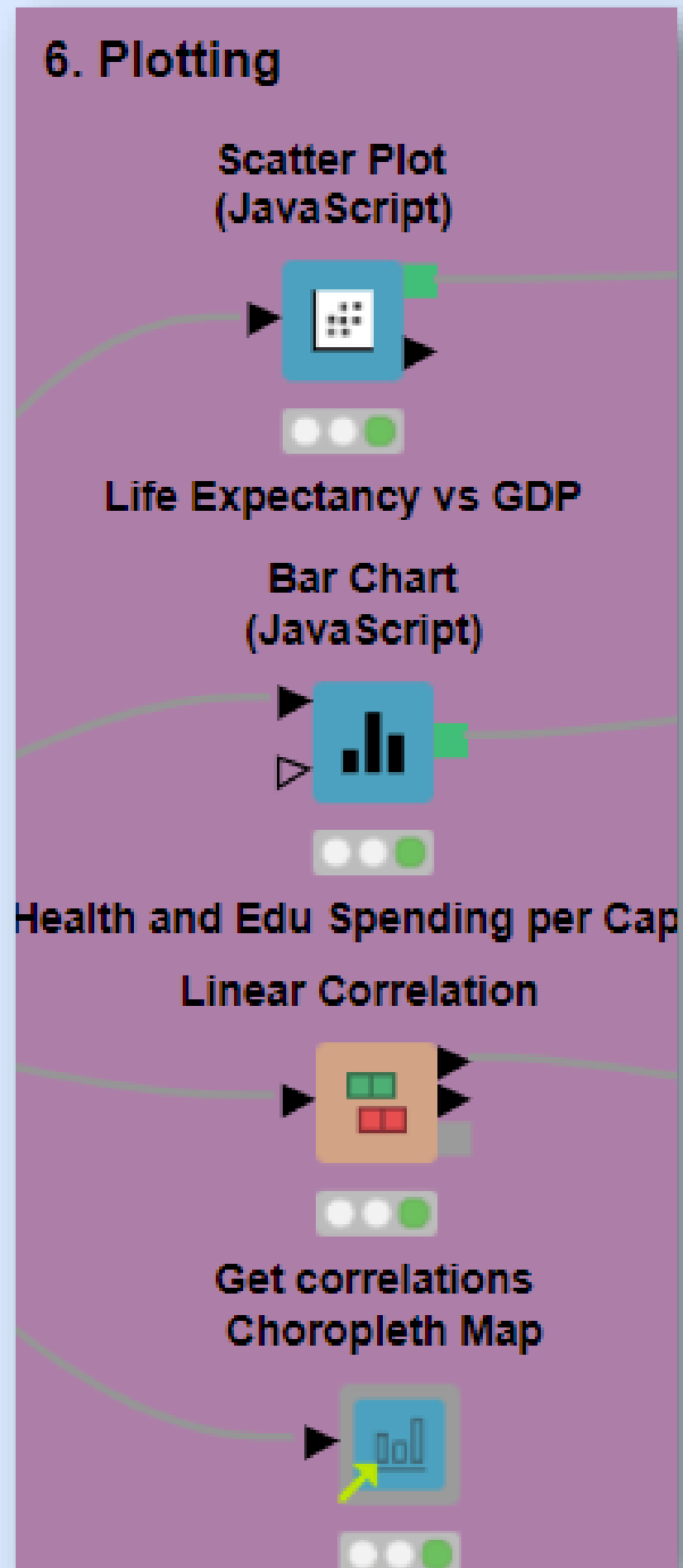


# Analytics and Visualizations

## Plots:

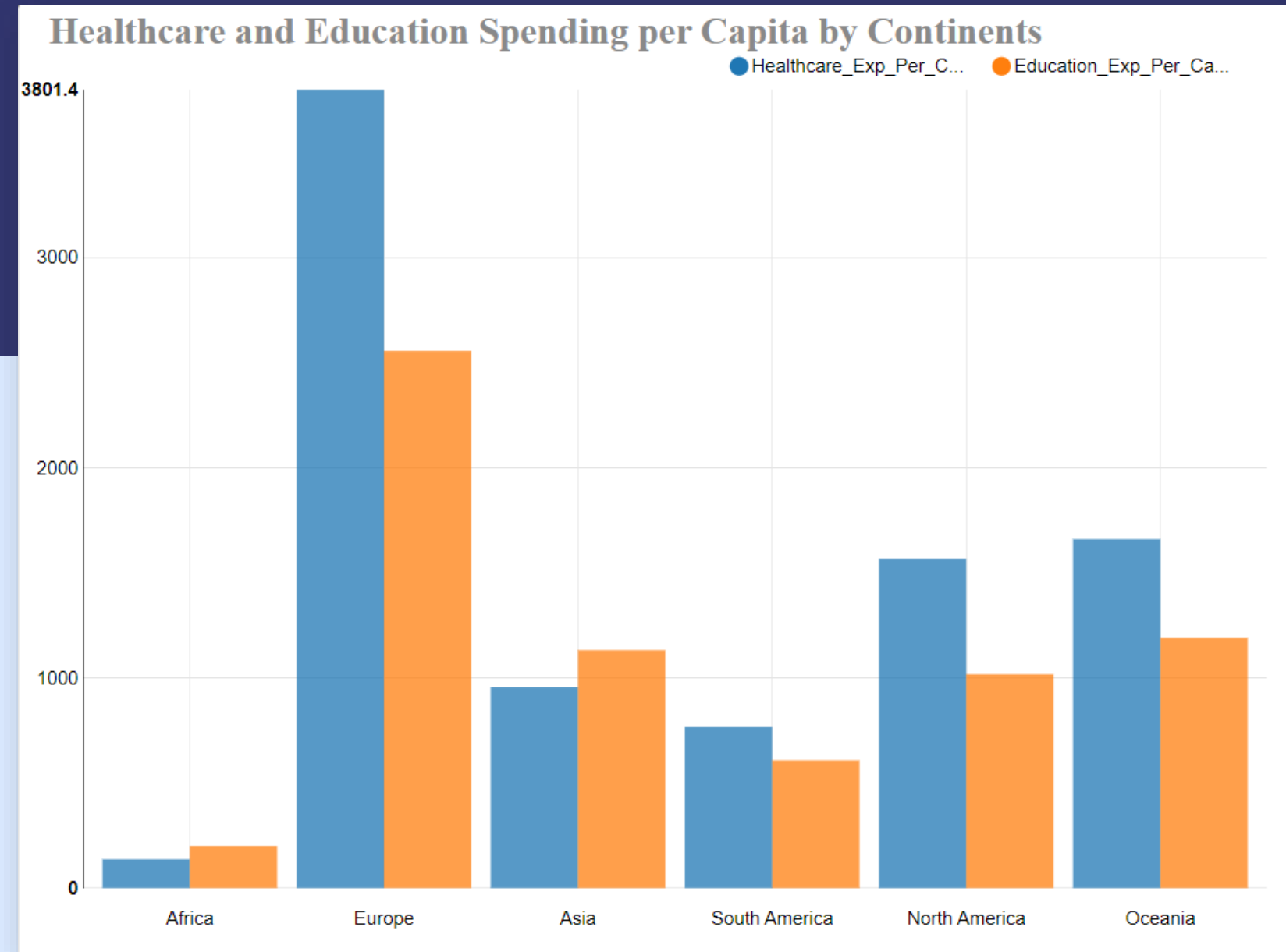
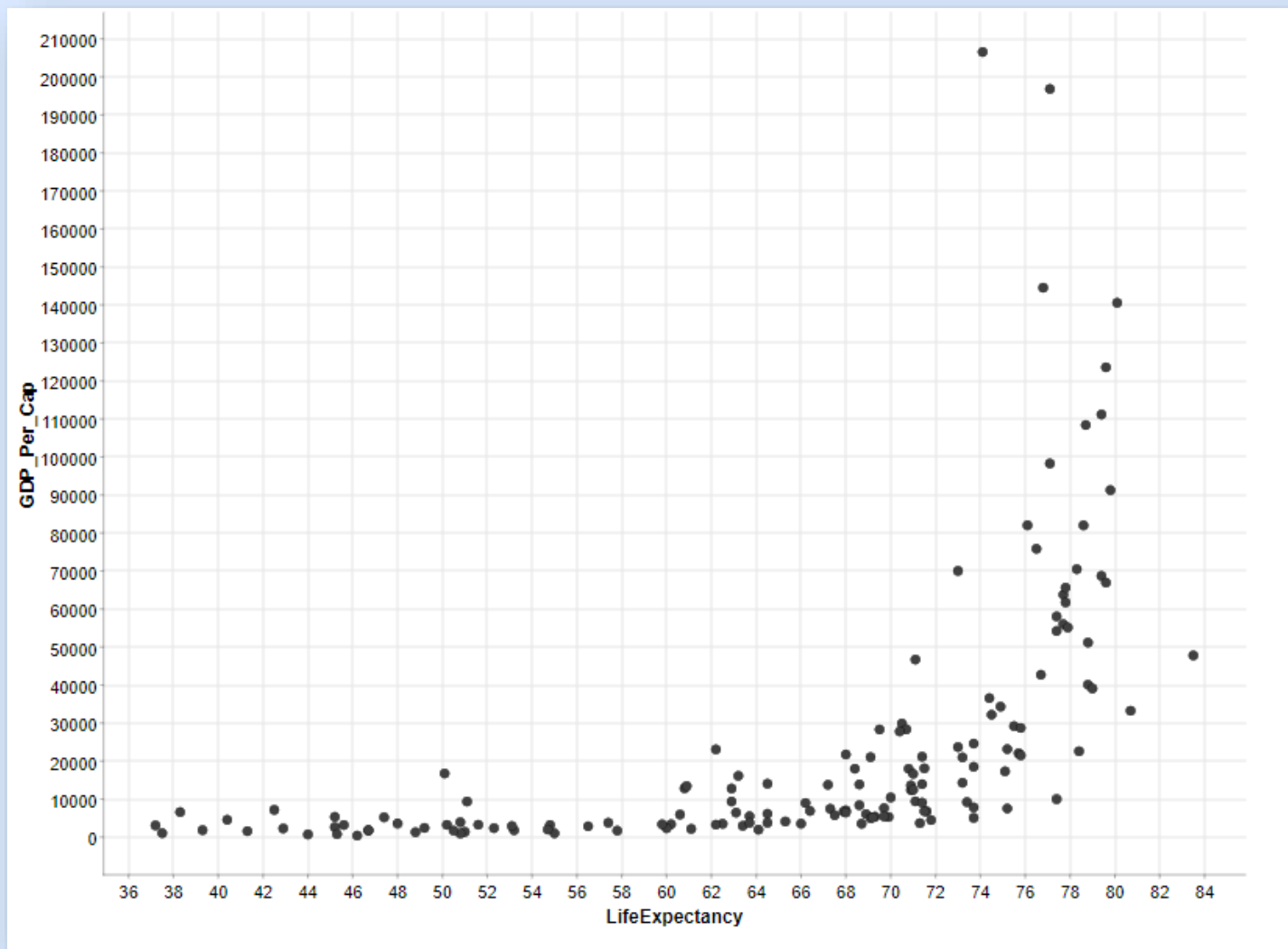
- Scatter plot: LE vs GDP
- Bar Chart: HE and ES

Analytics: Correlation matrix  
between our main variables





# Visuals

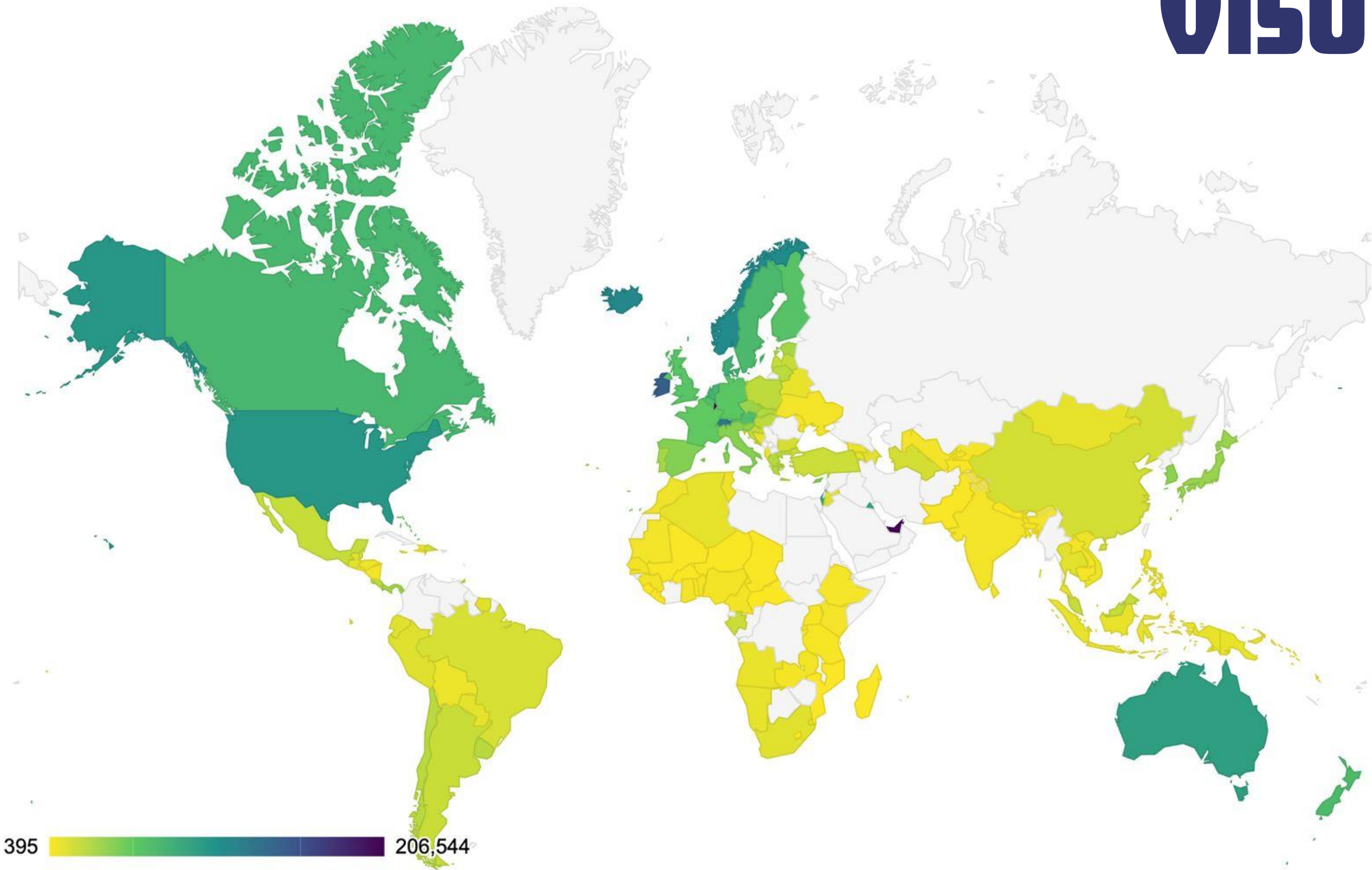


Row ID	LifeExpectancy	Healthcare_exp_per_cap	GDP_Per_Cap	Education_spending_per_cap
LifeExpectancy	1.0	0.578816739618267	0.5379352848154475	0.541396659994523
Healthcare_exp_per_cap	0.578816739618267	1.0	0.8006826249738677	0.8584553477637578
GDP_Per_Cap	0.5379352848154475	0.8006826249738677	1.0	0.9267381578254776
Education_spending_per_cap	0.541396659994523	0.8584553477637578	0.9267381578254776	1.0

GDP Per Capita

by Countries

Visuals



A series of ten horizontal bars of varying lengths in a light blue color, positioned on the left side of the image. The bars are arranged in a staggered, descending pattern from top to bottom, with the top bar being the longest and the bottom bar being the shortest.

**Thank you for listening!**