

Student name: Juweria Wajid Ali  
Course: MSc Data Science

# HISP Data Project

# Introduction

The project meets two of the requirements,

- the MSc in Data Science programme and
- is also a record of the work done during the placement at the University of Highlands and Islands.
- The goal of this project is to make a data visualisation of data about STEM student enrolments in the Highlands and Islands, Moray, and Perthshire. This will help the Highlands and Islands STEM Partnership (HISP) understand what students want to study, how the population is changing, plan activities, and predict what will happen in the future.

# Problem overview

The scope of this project would take into account the open data that is readily accessible, what it enlightens us with, and perhaps begin to look at the data gaps. This would allow to find out who has access to any non-open data that might be needed such as granularity regarding gender, deprivation, and rurality scores.

# Potential open datasets

## 1. Skills Development Scotland – Regional Skills Assessment(RSA) data matrix.

- Skills supply
- Skills demand
- Skills mismatch

## 2. Higher Education Statistics Agency (HESA)

- Who is studying in HE?
- Where do the HE students come from?
- Where do HE students study?

- What do HE students' study?
- What are HE students' progression rates and qualifications?

### 3. Scottish funding council (SFC)

- Proprietary data
- Name of FE UHI college, STEM subjects, Number of enrolments for academic years 2014-2015 to 2020-2021 for each UHI college by subject.

# Why cannot use all three together for this project?

- HESA: No statistics available specifically for the Highlands and Islands, Moray, and Perthshire regions. Also, there is no distinct definition of what constitutes a STEM subject. The HESA science category includes fields including agriculture, nursing, and medicine that may not be covered by other STEM definitions.
- SDS: Unable to release specific data related to the desired careers/preferred pathways, which were of particular importance to this project. This is because they contain personal level data.  
SDS offers funding to foundation apprentices who study via local authorities (councils).
- SFC: Funds and counts students attending colleges and universities.  
SFC do not monitor apprentices outside their funding.

# Scottish Index of Multiple Deprivation(SIMD)

- ➔ A comparative indicator of deprivation across 6,976 local regions is the Scottish Index of Multiple Deprivation (called data zones)
- ➔ Seven categories examined are income, employment, education, health, access to services, crime and housing.
- ➔ If a region is labelled as "deprived," this may refer to people having poor incomes, but it may also indicate that there are less resources or possibilities available.

## Correlations SIMD vs STEM

- ➔ While it made sense with the SIMD to use the 2016 and 2020 data, however, the STEM data is available for academic years i.e., 2015–2016, 2016–2017, 2019–2020, and 2020–2021. Some data for the year 2016 will be available in the academic year 2015–2016 and 2016–2017, the same for year 2020. Choosing STEM enrolment data was unclear in order to look for correlations.
- ➔ The SFC information is collected annually and tends to be available in January for the preceding academic year. So, in January this year i.e., 2022 the data was available for academic year 2020-21. An assumption was made to select January 2016 and January 2020 STEM enrolments for correlation plots.

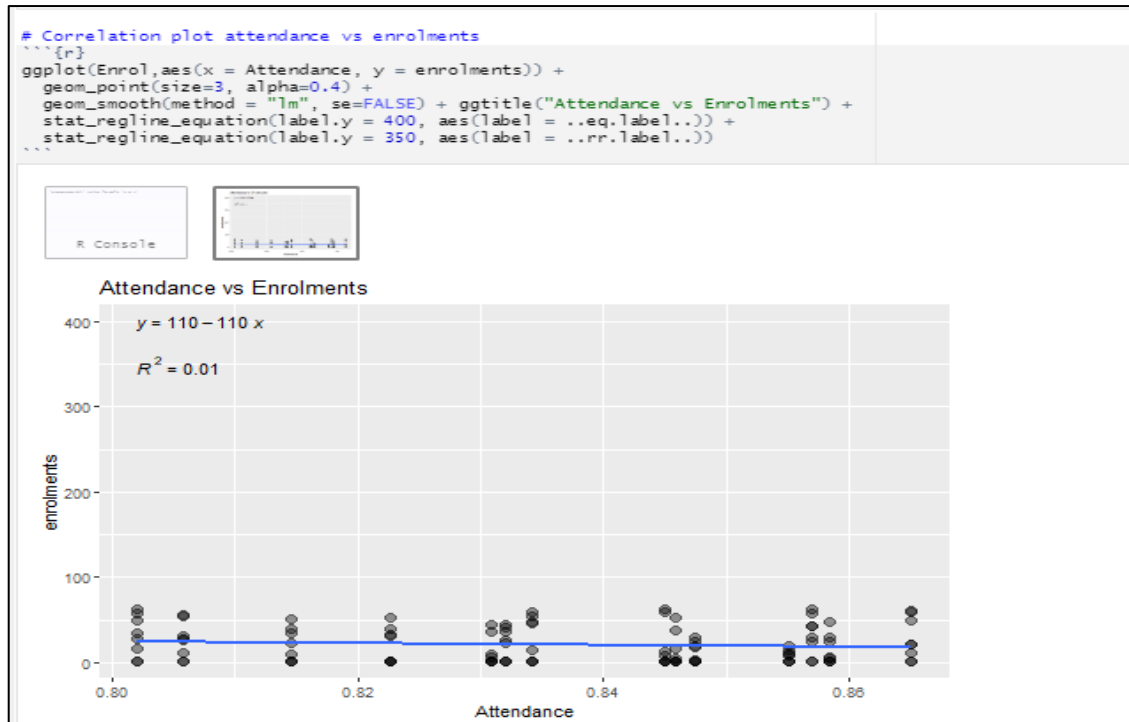
- Attendance vs Enrolments

Data: SIMD 2016

SIMD 2020

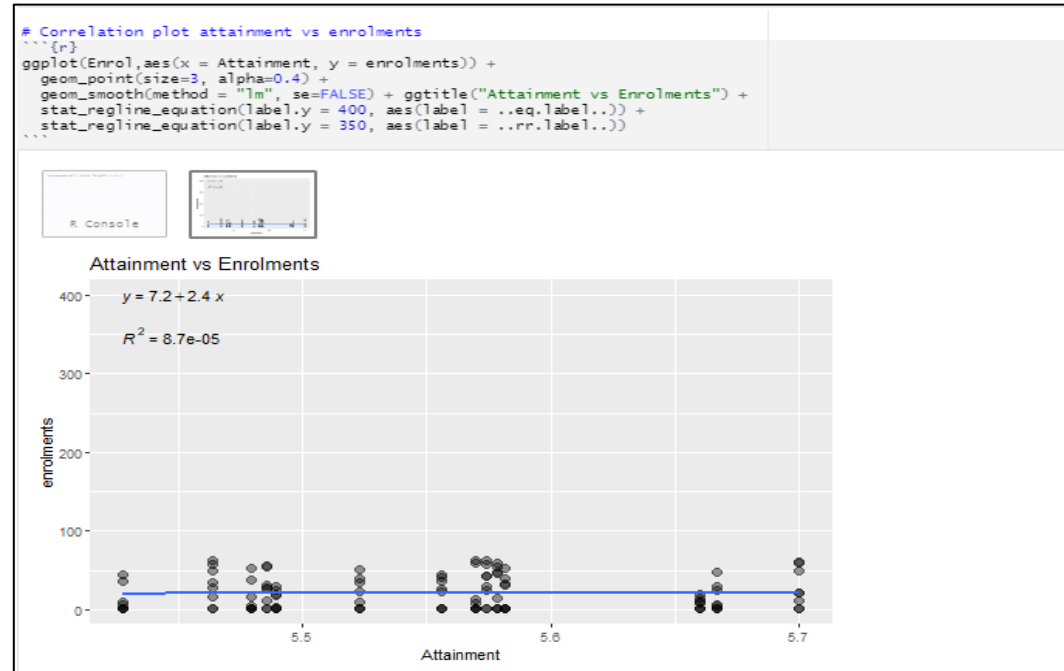
SFC Jan 2016

SFC Jan 2020

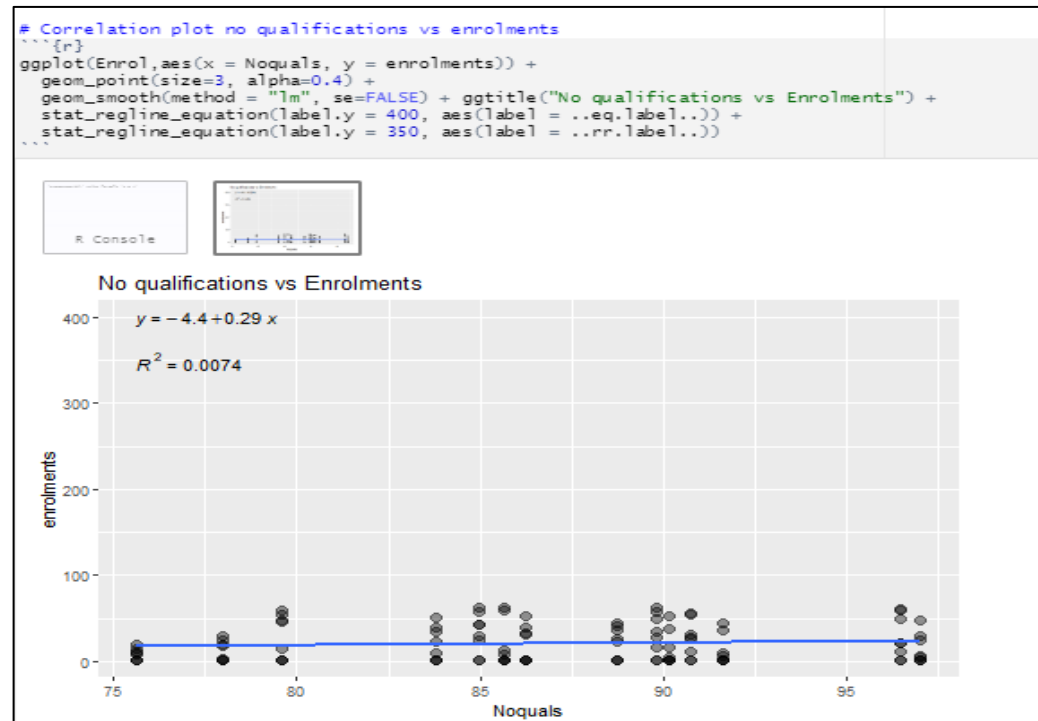




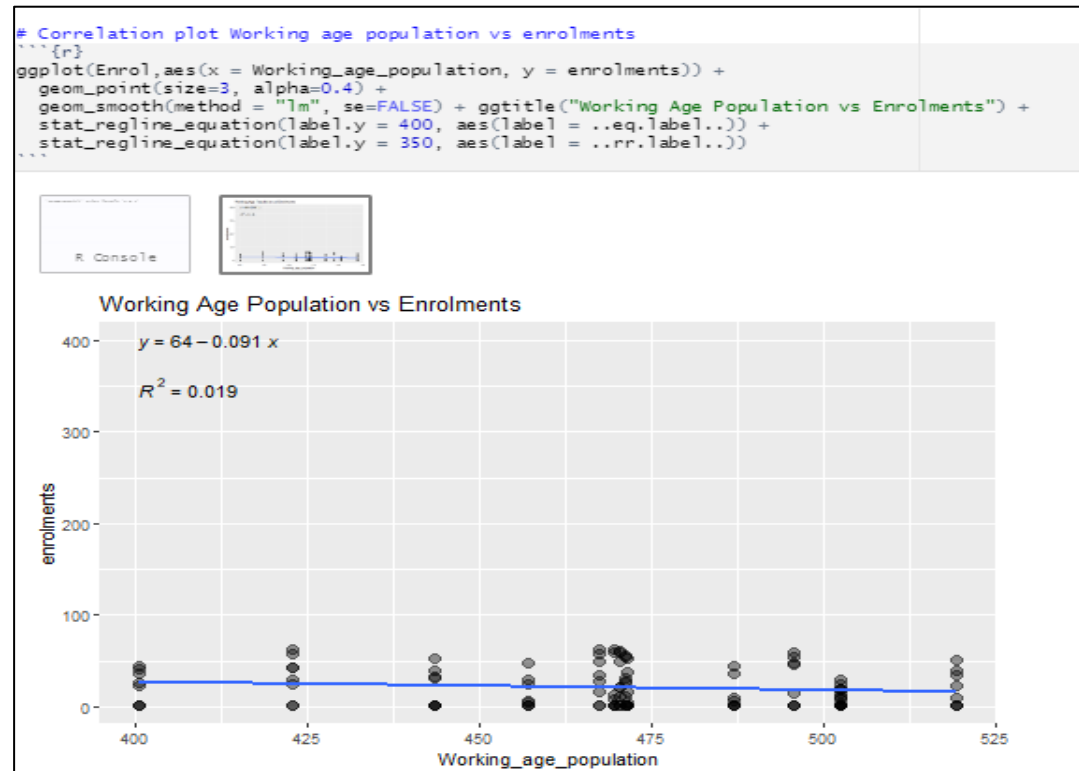
- Attainment vs Enrolments



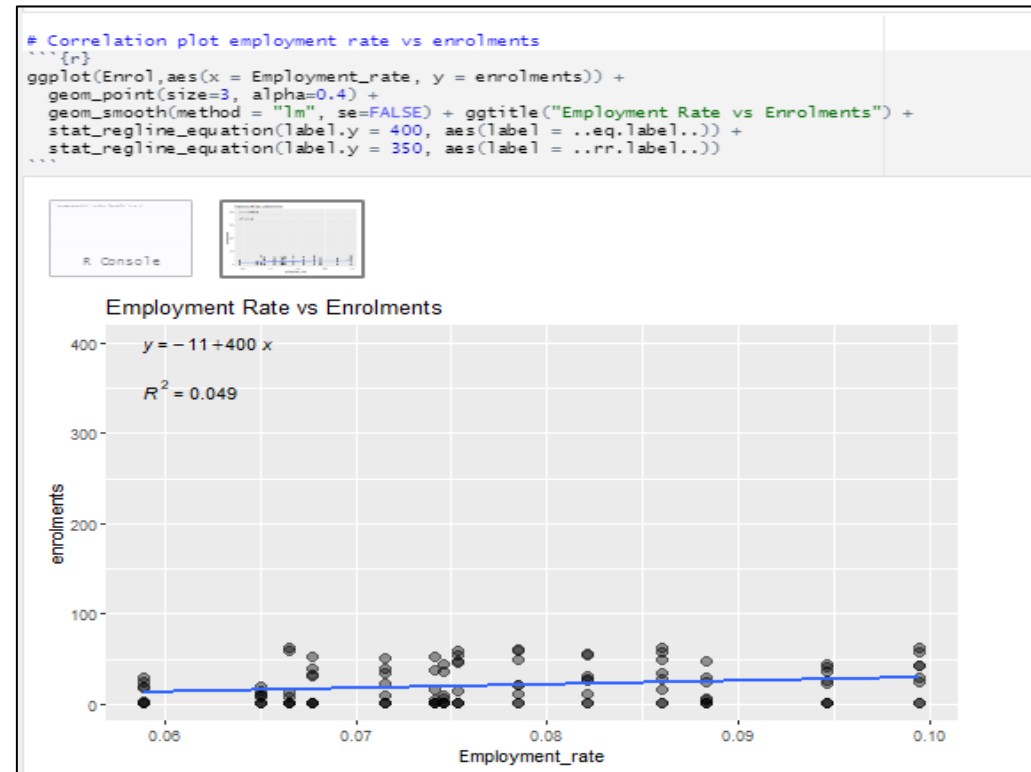
- No qualifications vs Enrolments



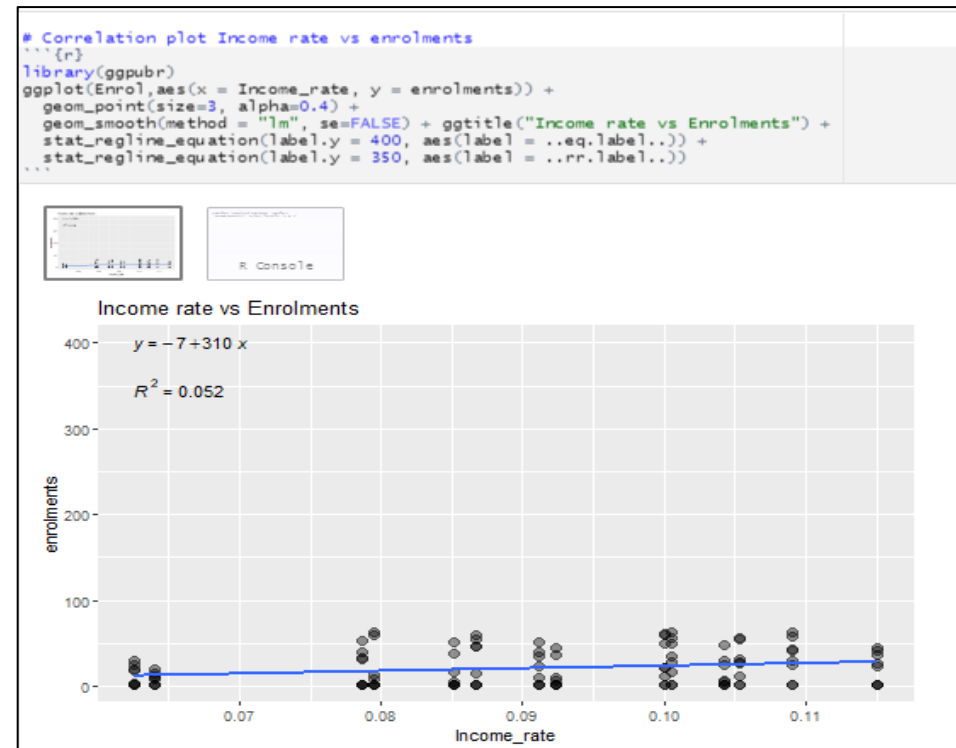
- Working age population vs Enrolments



- Employment rate vs Enrolments



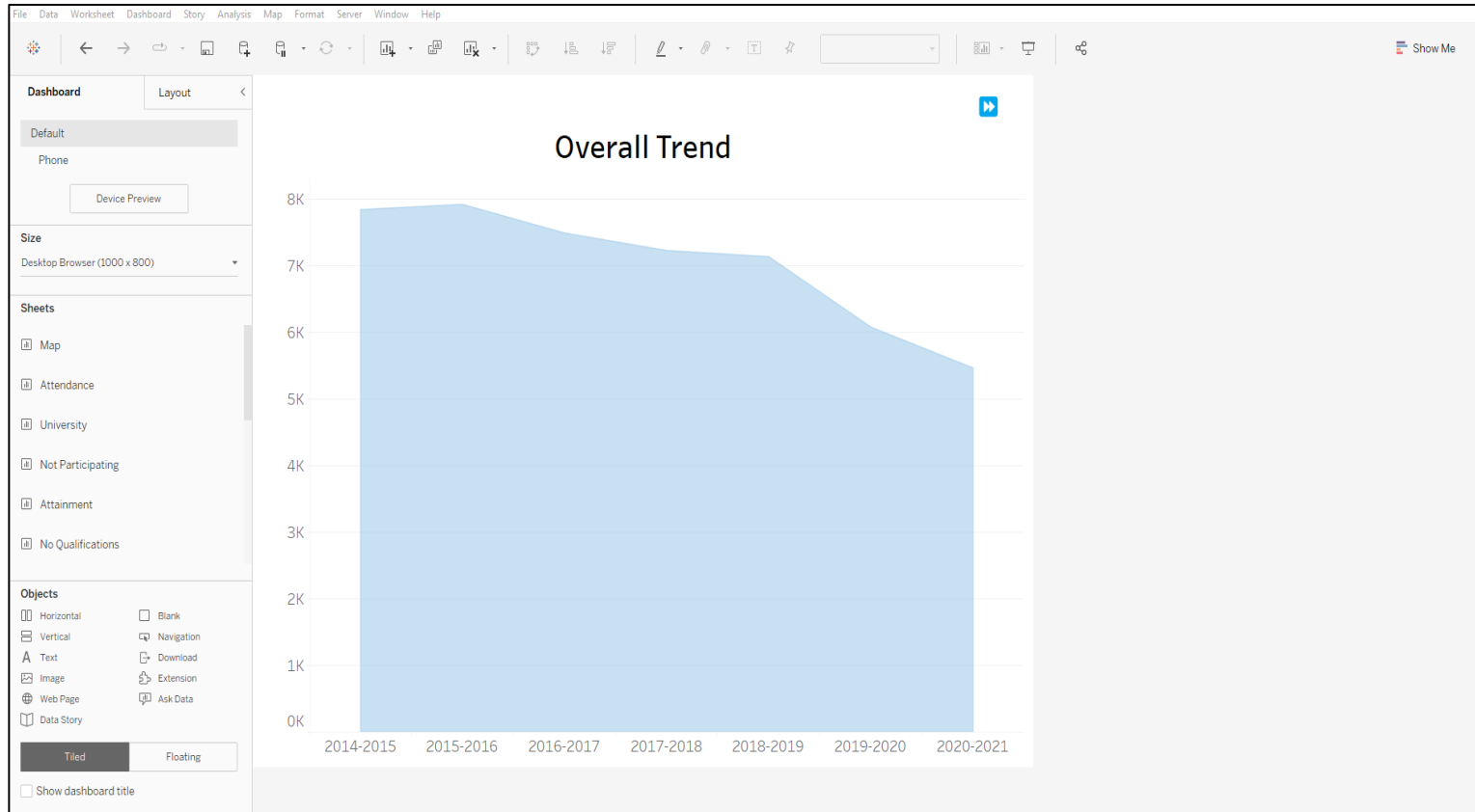
- Income rate vs Enrolments



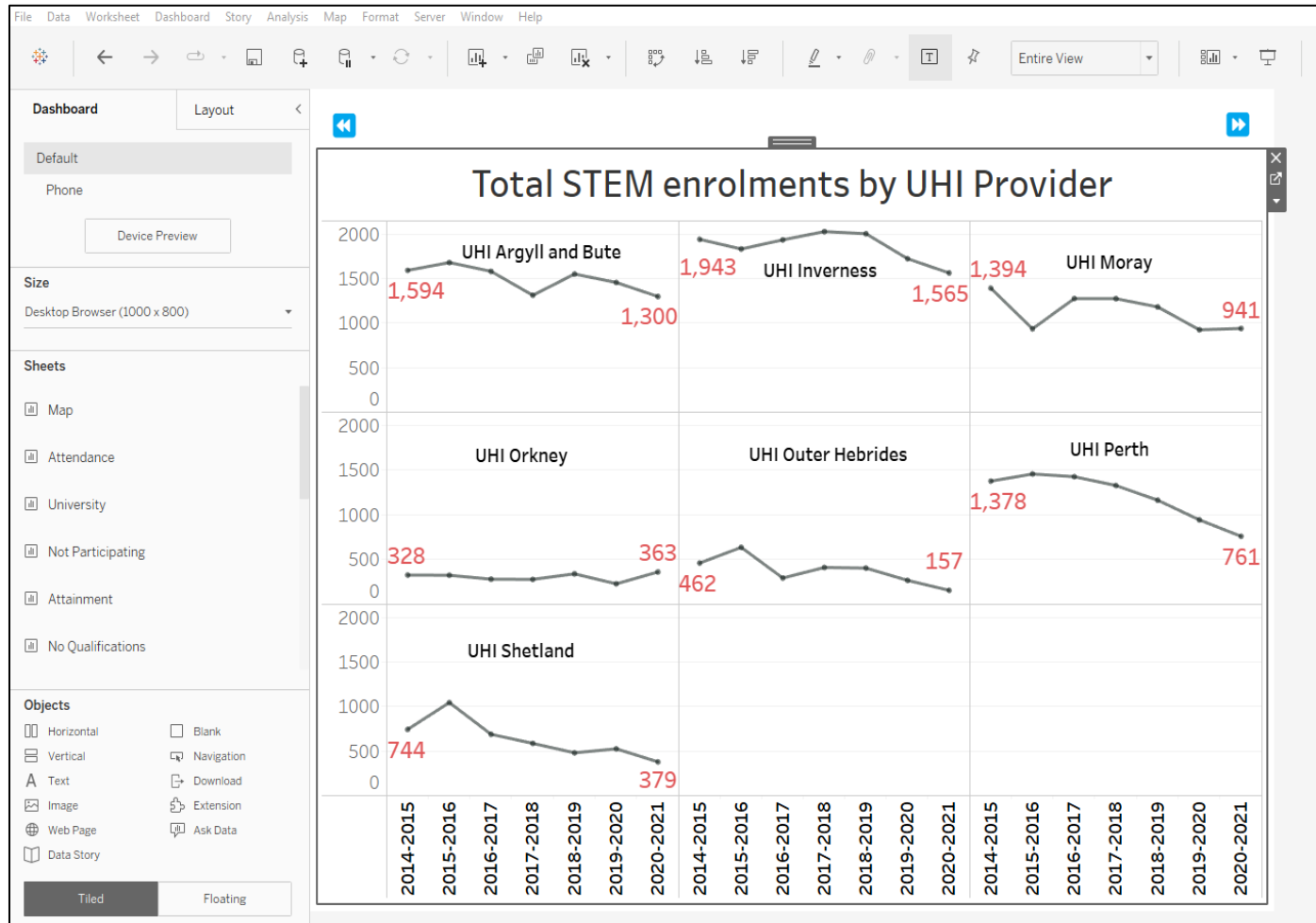
# **STEM Dashboard**

- Tableau Desktop has been used to build these dashboards
- STEM FE Colleges in UHI region
- No data available for UHI North Highland and UHI West Highland
- HE data not available.

- Overall Trend

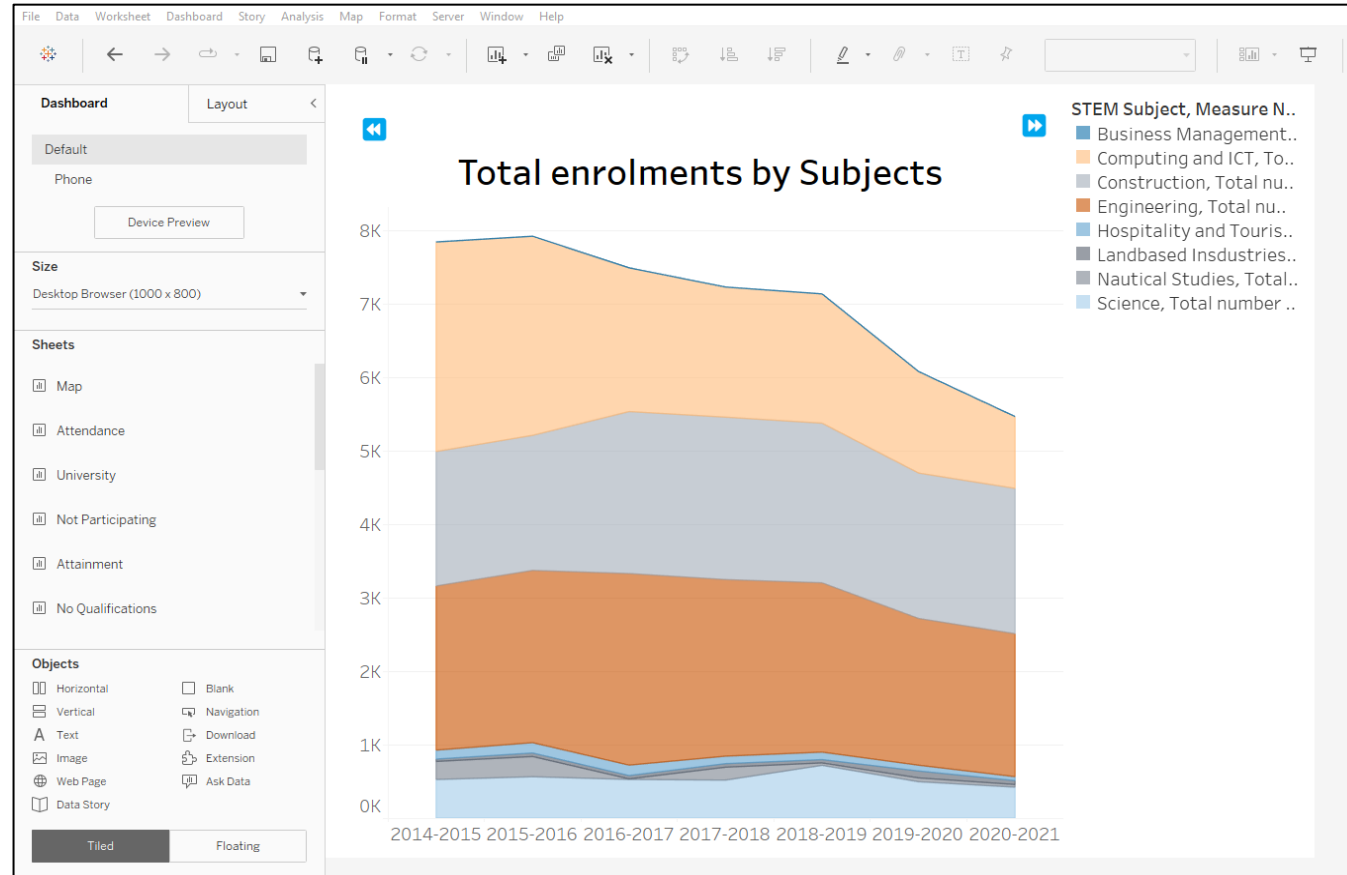


- Trend – STEM enrolments by UHI College

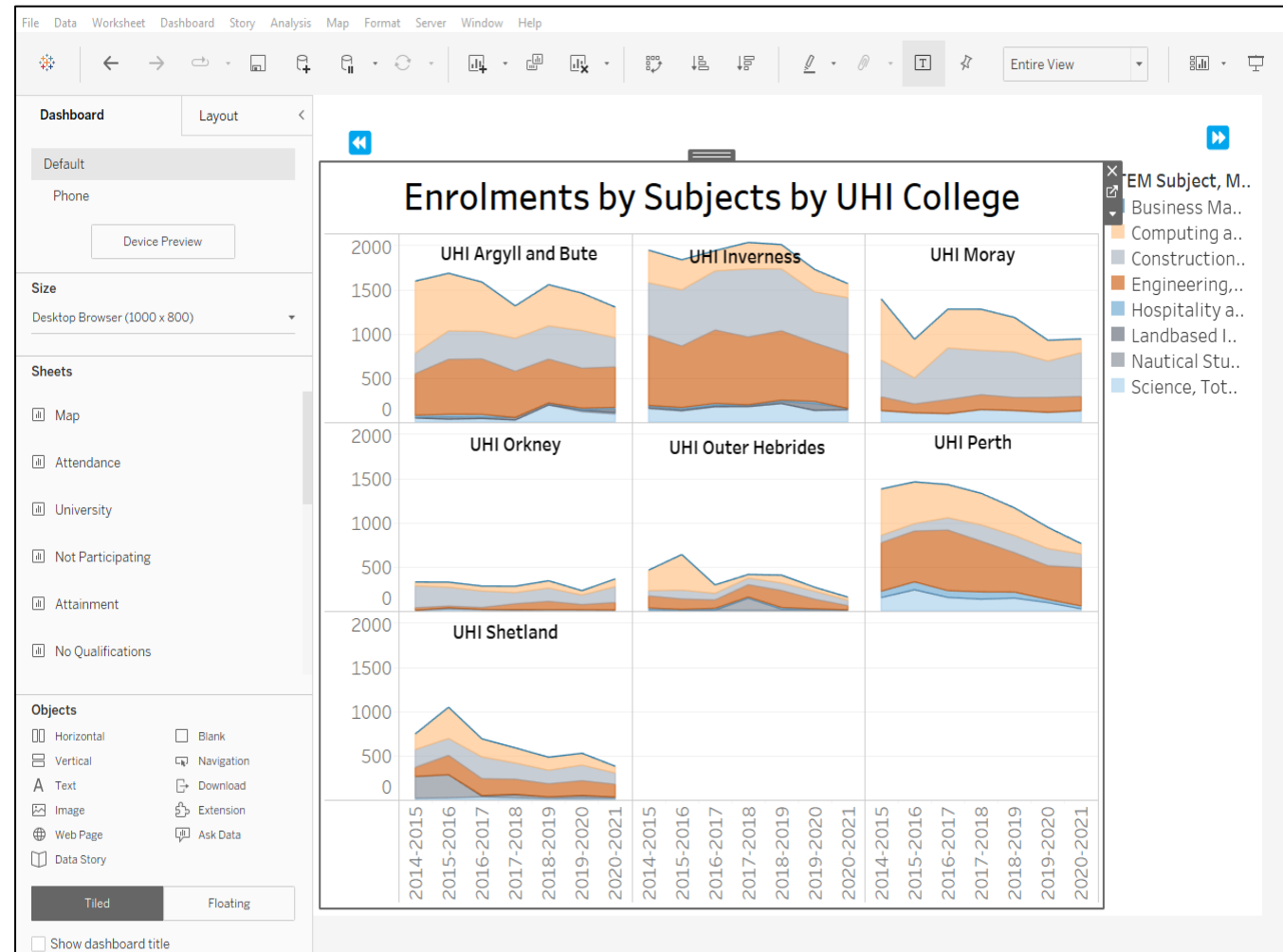




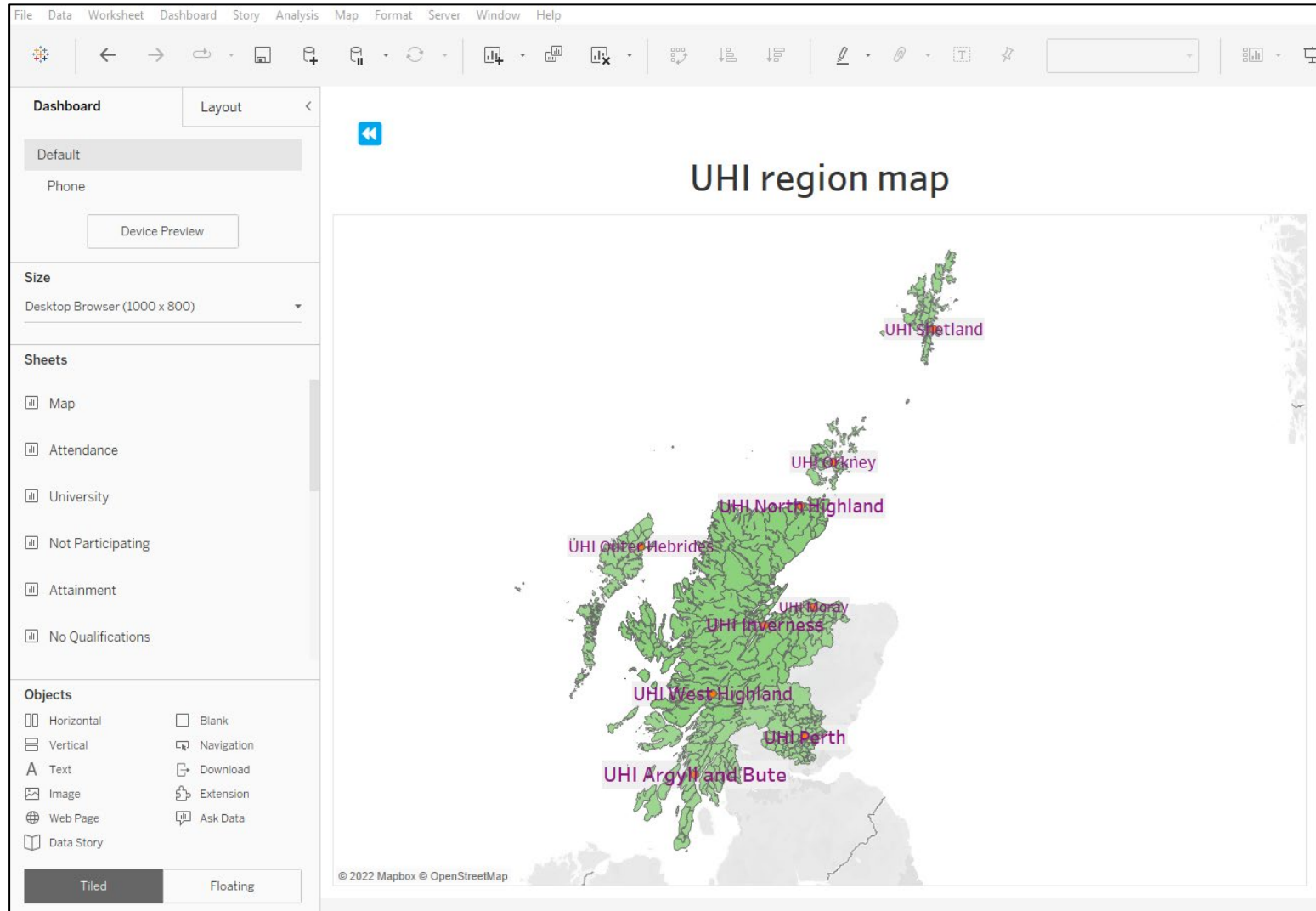
- STEM enrolments by Subjects



- STEM enrolments by Subjects by UHI College



- UHI Region Map



# Future Scope

- SDS, HESA and SFC will not be able to be put to use together for this project because, despite the fact that each of them, in its own way provides very insightful information, there is no commonality between them that could be used to gain insights. These could be utilised in future projects if common grounds can be identified.
- UHI North Highland and UHI West Highland data can be added once available.
- HE data once available can be used in a similar way to gain insights.
- At this point in time there are not enough data points to perform predictive analysis, hence no machine learning algorithms could be used. This could be added to the future scope.
- Effect of COVID-19 on the enrolments could be identified.
- User testing/usability – Although developmental testing has been carried out when building the dashboards.

## Conclusion/Recommendations

- This is an exploratory Data Science project
- The outcome of this project can be used to determine future directions in data collection and sharing, access, and improved planning for the promotion of STEM professions.
- Tableau Reader is a free application that can be used to open and see workbooks that have been built in Tableau Desktop. Workbook saved as a .twbx and its data source can be emailed to others, and they can open and see its contents using Tableau Reader.

<https://www.tableau.com/products/reader>

Thank You