B209415 discriptive report

Exploratory analysis

Data used for this report is part of the accidents and emergency attendances and admissions data which is a part of NHISdatasets package. As accidents and emergency dataset is very large with 12,765 raws, it was divided into training (which was used over the course) and test data (which was used to collect data through the data capture tool in python). To generate the test data an index column was added to the accidents and emergency data. Then the needed variables has been selected (here index, period, attendances and breaches was selected to be included in the test dataset and in the data capture tool, then generating the proportion of data as training data (the training dataset will have the majority of the data and it is generated randomly). Finally the test dataset is the subset of accidents and emergency data after subtracting the training data. The final accidents and emergency attendances and breaches test dataset that will be used for data collection is a CSV data with 12 observations and 4 variables. The variables are: index (an integer variable that give sequential number for each observation), period (date variable (integer), indicate the month in which each activity happened), attendances (decimal variable, denotes the number of attendances in each month) and breaches (decimal variable, denotes the number of attendances that breached the four hour target). Information on the average attendances and breaches over years will help policymakers monitoring the progression of hospitals performance over the whole England.

index	period	attendances	breaches
930	Jan-17	8,609	2,619.0
2009	Oct-16	20,345	1,047.0
3181	Jun-16	1,590	0.0
5148	Jan-18	357	0.0
5229	Jan-18	4,922	24.0
6104	Oct-17	1,169	0.0
6700	Aug-17	17,059	1,640.0
8146	Apr-17	3,848	0.0
8479	Mar-19	506	0.0
10524	Oct-18	7,848	1,343.0
11428	Jul-18	435	0.0
11655	Jun-18	876	4.0

Analysis

The data generated through the data capture tool (the tool was generated in python) is a CSV file and called createdData. The data entered from accidents and emergency test data. The tool captures five variables (index, period, attendances, breaches and consent). So, it contains the same variable in the test data except the consent which is added in the tool and is extremely important to protect the ethical values while collecting the data, and it is a logical variable. Various types of exploratory and explanatory analysis can be done based on this data which will be valuable information for policy makers. It allows us to know the total, average, minimum and maximum attendances and breaches across the England hospitals over the period of four years. It will also allows measuring the differences and comparing attendances and breaches for all England in different years. This information is valuable as it will show which year had the peak of attencendeces and breaches and whether this difference is statistically significant compared to other years. This may motivate further investigations on the cases of this peaks and differences. According to the data that was collected by the data capture tool, though 2018 had the highest average attendances it had the lowest breaches. 2019 had

the lowest average attendance. Further analysis is illustrated in tables A to D and the graph.

Descriptive statstics

TableA:summerize the attendences data

meanattendances	SDattendances	minattendances	maxattendances
7744.917	5157.915	334	17247

TableB: summerize the breaches data

meanbreaches	SDbreaches	minbreaches	maxbreaches
632	821.3927	0	1977

TableC: Showing the summary of attendences and breaches by year

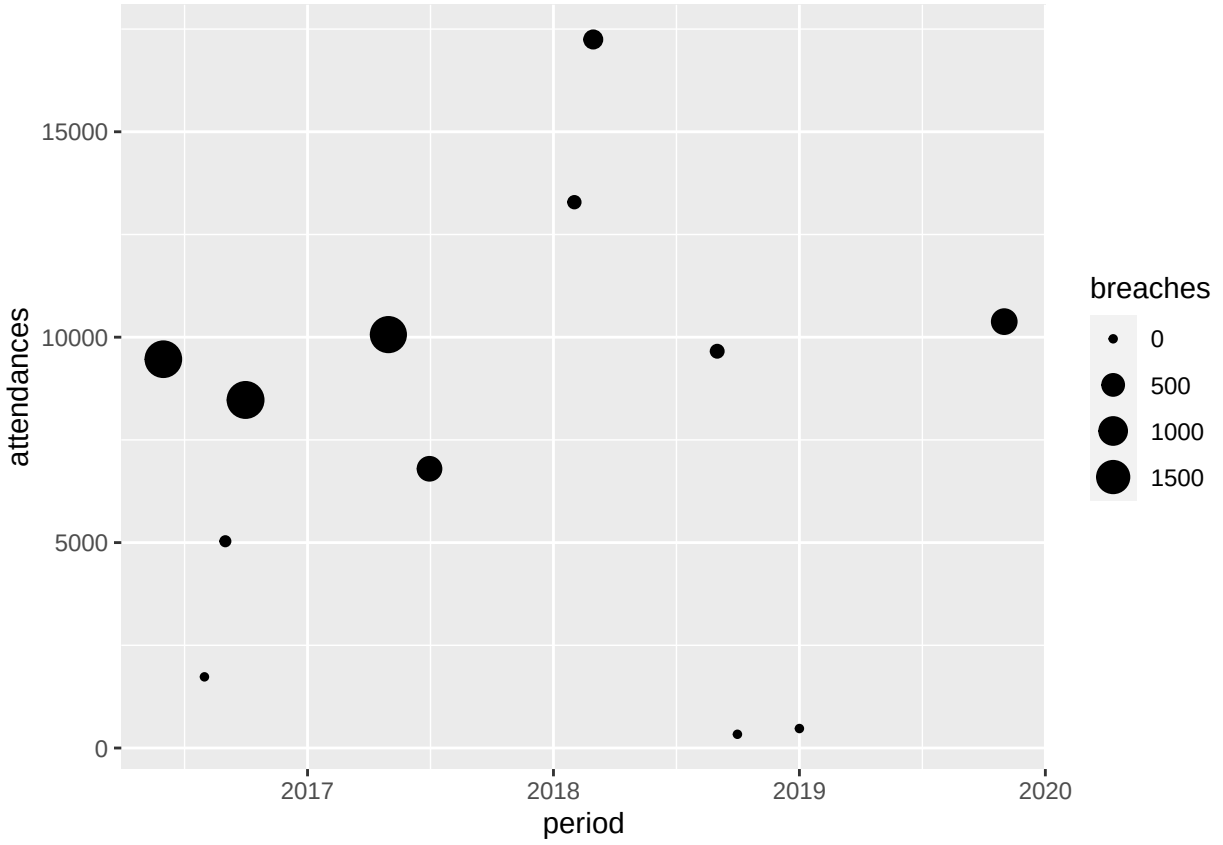
period	meanattendances	SDattendances	sumattendences	minattendances	maxattendances
16	6175.25	3518.434	24701	1732	9464
17	8430.00	2309.411	16860	6797	10063
18	10131.25	7229.486	40525	334	17247
19	5426.50	7003.893	10853	474	10379

TableD: shows the summary statstics for breaches by year

period	meanbreaches	SDbreaches	sumbreaches	minbreaches	maxbreaches
16	983.75	1126.2390	3935	0	1977
17	1259.00	869.7413	2518	644	1874
18	101.00	118.8697	404	0	273
19	363.50	514.0666	727	0	727

Creation of graphs that shows the the attendances and breaches by year

The period is in the x-axis, attendences in the y-axis, whie the size of the puple represent the number of breaches amonge attendences each year



Data Management

ADMINISTRATIVE INFORMATION

Institute: Deanery of Molecular, Genetic and Population Health Sciences, University of Edinburgh Supervisor: Dr Mairead Bermingham, mairead.bermingham@ed.ac.uk

Project start - end date

2022-04-06 - 2022-04-07

Data Collection

Data collection tool was generated in python to collect part of accidents and emergency attendances and breach data provided by NHIS datasets. The data collection tool capture 5 variables: 1- Index. 2- Period. 3- Attendance. 4- Breaches. 5- Consent. The data used for data collection and the data generated through the data capture tool are both in CSV format.

Documentation & Metadata

The data collected by the data capture tool and variables were described and metadata was created using the dictionary function within R. Then the dictionary was saved in a separate file within the data folder.

Ethics & Legal Compliance

The project will comply with the college of Arts, Humanities and Social Sciences (CAHSS) research ethics framework and guidelines at the university of Edinburgh

STORAGE AND BACK-UP

The data will be organised into folders. Each folder for specific type of data and then the data and the changes made to it will be stored in the Notable and the Github repository of the principal investigator

SELECTION AND PRESERVATION

Long term data storage The data will be kept in an online repository for one month to give the collaborator an opportunity to assess the project after it is finished. After that the data will be kept in the University of Edinburgh DataShare repository. The data that will be stored The original data and all the data derived from it, the r codes and python scripts that has been used to import, manipulate, collect and analyse the data as well as the results of the analysis.

DATA SHARING

After obtaining the ethical approval from the The data will be shared through the NHS the data will be share as follow: Throughout the project 1- Git hub repository 2- The university of Edinburgh SharePoint After the project 1- University of Edinburgh DataShare data repository

RESPONSIBILITIES & RESOURCES

The principal investigator is responsible for data management of this project Further training of how to upload to the university of Edinburgh SharePoint and Data share is required.

Coding

Start with a new project and clean environment, this will help in putting all activities in one place and avoid confusion in folders and errors. Writing the project name, the author and the link to Github in the header of the R markdown file. Installing and loading all necessary packages needed for the various codes that will be used. To ensure that I have effective codes i followed the following practices: Commenting and annotating before each code: This will help in code readability as it will allow anyone who reads through the document to understand what has been done and the intention of each code. Avoiding writing large chunks of code: Also to improve the readability cplex and long coding was avoided rather the code is divided and explained and steps were declared as (firstly, secondly, thirdly,. . .). All the codes were structured in a way that allows having the same results whenever it is run. To ensure this version control through Github repository was used both to save the data upon which the code depends and to allow tracking of changes to the codes and data. Both of the above mentioned practices will ensure code availability and reliability. Storing original files, the codes, data produced in different folders but in the same directory. Using meaningful names: When assigning a name to data the name given is related to what data have such as accidents and emergency test data (ae - Test) or abbreviation of the data name (CreatedData: CD). This will help in avoiding unnecessary confusion. To avoid unnecessary error most of the codes used is a modification of already existing codes from the course examples. Closing of the project and running all the codes from the start to ensure that all codes are functioning whenever they called (ensuring availability and reliability) Saving correctly all the folders and files. Commenting and pushing all the changes made to the Github repository

Data storytelling

The goal is to know the situation of attendances and breaches in accidents and emergencies in England hospitals over the years and the difference in these figures. My target audience are the policy makers in the health sectors, as they guide implementation of various practices. This information will help them to

understand the effect of activities in various years. It may also help them in investigating more in this topic. The data capture tool records the date, the attendees and breaches. The data resulting from this tool can be analysed in a very simple way (means/year. totals/year and the average across the year) and then this data can be presented in simple tables and graphs.

Future work

Next time I will be working with my tool I will start by detailing my data management plan, and how I am going to organise the folders within Notable and Github. Then I will further communicate with my collaborators/supervisors and motivate them to revise and edit the tool and use the version control in Github. Then I will try the tool at different times to make sure all codes are working effectively. I will also make sure that all the codes are written according to the best practices.

Reflective practice

#Data management process For me the data management process was the most challenging part especially using DMPonline. This has led to many troubles in the codes and data I used for the first assignment. However, after in depth studying about it from the course content, the discussions raised by my colleagues, and some external sources, I revised all the work I did previously and I was able to handle and analyse the data in a much more useful, and reproducible way. #Coding practice As I have already studied the coding in R and python in the previous courses, I was familiar with most of the coding practices and I was trying to apply it throughout my coding. From readability, to availability and reliability.However, reliability was a significant problem for me as many codes were not running correctly when I was trying to run them later, so here I found the problem was basically in my data management process and not the code itself here I changed the way i handled the data (in terms of storge, version control..) and then I was able to write a good quality codes. #Response to feedback Though I used to have very poor input in the discussions posts especially in the first two weeks of the course. The discussions posts and my colleagues' thoughts, questions and responses to my post have significantly shaped my skills in coding and handling data. Many coding, data management and dictionary development difficulties that I or my peers posted we found timely and efficient solutions that I applied to my practices. #Skills developed Throughout this course I learned many new skills, such as building the GIThub repository and how to connect it to my projects in R and python and then keeping track of all changes I made through the version control. Also, I learned how to build a data management plan utilising the DMPonline and then how to apply the plan for the data. Furthermore, I learned how to build a data capture tool in python. This course significantly improved my critical thinking and problem solving skills, as I learned many different ways of handling code errors and the importance of trying to understand the error itself. Finally, if there is one thing missing is the communication skills as I was not effectively participating in the discussions board and I missed a huge oppertunity.