

# Public Health Scotland Group Project - CodeClan, DR-16

## Names of group members

Lloyd Lombardi  
Jamie Zike  
Nico Mastronardi  
Neil Plenderleith

## Roles & responsibilities of each member

We were all involved in every process to some extent but the main priorities and responsibilities were as follows:

### **Lloyd** worked on ...

Data Investigation, Cleaning & Wrangling  
Visualisations  
Excalidraw wire framing for dashboard  
Dashboard

### **Jamie** worked on ...

Data Investigation, Cleaning & Wrangling  
Visualisations  
Excalidraw wire framing for dashboard  
Breakdown of the brief  
Presentation

### **Nico** worked on ...

Data Investigation, Cleaning & Wrangling  
Visualisations  
Trello breakdown  
Geo Spatial with sf and leaflet

### **Neil** worked on ...

Data Investigation, Cleaning & Wrangling  
Visualisations  
Excalidraw wire framing for dashboard  
Geo Spatial with sf and leaflet

## Brief description of dashboard topic

Our topic is to what extent is the NHS “winter crisis” real? How has winter impacted NHS Scotland in the past? Why might it be even more critical in the coming year and might the pandemic influence the potential scenario?

The dashboard outlines our topic by examining these main questions and the numerous other sub questions in the brief regarding demographics and how these might drive hospital activity.

### **Stages of the project**

- \* Data Investigation - ideas posted to shared Padlet
- \* Question breakdown
- \* Trello workflow
- \* Visualisation for patterns
- \* Dashboard wireframes on Excalidraw
- \* Geospatial with sf and leaflet
- \* Git branching & version control
- \* Dashboard construction - pulling together of everyone's graphs and ideas
- \* Presentation and documentation

### **Which tools were used in the project**

- \* Zoom (daily stand-ups and occasional mob programming)
- \* Trello (planning & task allocation)
- \* Git/GitHub (collaboration & version control)
- \* Padlet (visualisation ideas and record keeping)
- \* Excalidraw (wireframing for dashboard)
- \* Slack (messaging and code sharing)
- \* Google Slides (presentation)
- \* R, R Studio and R Shiny

### **How did you gather and synthesise requirements for the project?**

This was achieved by drawing out the main questions of the brief, breaking these down into more manageable sub-queries in a shared google sheets document. We then began investigating the data with visualisations to draw out trends and patterns pertaining to these sub-questions. This became an iterative process of data investigation, visualising from the data and linking back to the brief in order to answer the sub questions.

### **Motivations for using the data you have chosen**

We pulled data from every dataset in the end to answer our subqueries. This is reflected in our dashboard which contains multiple tabs, each focusing on one particular demographic or sub query from the brief.

### **Data quality and potential bias, including a brief summary of data cleaning and transformations**

- Data quality

According to the About tab on the dataset page/dedicated page online, the data quality meets the highest standards of trustworthiness, quality and public value. This publication fully complies with the Code of Practice for Official Statistics. Further information can be found here:

<https://www.publichealthscotland.scot/publications/acute-hospital-activity-and-nhs-beds-information-quarterly/acute-hospital-activity-and-nhs-beds-information-quarterly-quarter-ending-31-december-2021/data-quality/>

- Data cleaning

The datasets were relatively clean for PHS Scotland. We joined a health board names dataset on any health board datasets in order to add in the Health Board names as these were omitted in favour of the Health Board code, this allowed us to more easily compare metrics by health board.

We had to do some transformation around the dates of the data as some datasets were recorded month by month and others by quarter. In the Covid datasets, we also added a logical variable to compare winter (Dec/Jan/Feb) and non-winter months.

Shapefiles for the geospatial plots were imported from Public Health Scotland.

The A&E dataset had to be transformed into long format to split up attendances at A&E by destination when making graphs. Additional, in order to explore A&E attendance and destinations during Covid and compare with pre-Covid averages, we extracted comparison means for attendance and destination proportions from 2017-2019 from the dataset and added them to a Covid A&E dataset, which was used to construct the graphs in the dashboard tab that focused on the impact of Covid.

- Data storage and structure

NHS data is stored as linked data; linked data is that it is “data that can be linked to”. In practice, this means that wherever possible, any data point contained in a linked dataset should have a unique identifier which is a URL. Doing this means that the data point can be browsed to using a web browser (if a human-readable web-page has been created). It also means that the data point can be referenced by other linked data sets. Further info can be found here: <https://guides.statistics.gov.scot/article/34-understanding-the-data-structure>

This makes it very powerful for connecting datasets, and for providing links to data sources.

- Ethical considerations, bias and data use

There was no indication that the information in the datasets themselves was biased and the accuracy blurb at the dataset information indicated that PHS receives data directly from the health boards, undertakes a quality check and compares it to previously published data and trends. However, there may be ethical considerations, because there is potentially important information missing from the datasets, which could in turn have caused us to overlook some important effects and biased our analysis.

For example, we don't have information on staffing levels during the pandemic, which could have helped make more sense of the winter crisis and the impact of Covid-19 on the crisis. Additionally, we don't have demographic information about race or self-reported disability, so even though we know from other research that minority ethnic groups and disabled people were disproportionately impacted by Covid-19, this isn't included in the datasets and thus the effect isn't visible in the analysis. Lastly, we didn't have data on deaths,, so we weren't able to look at the differences between winter deaths during Covid-19 vs pre-Covid-19. This means our conclusions in the analysis could be biased.

The datasets are covered by the Open Government License, which means that anyone is free to use the information contained in the datasets. Specifically, anyone can copy, publish, distribute, transmit and adapt the information as long as they acknowledge the source of the information in the product or application. Where possible, users of the information should provide a link to the licence.