

# Combined R Markdown document

B210741

18/06/2022

## 1. Link to Github repository:

[https://github.com/B210741/B210741\\_assessment.git](https://github.com/B210741/B210741_assessment.git)

## 2. Constructing a data dictionary and appending it to the data

### Loading NHSRdatasets

```
install.packages('dataMeta') library(dataMeta) library(tidyverse) library(here)
```

### Data

The data you will be managing on the course are from the NHSRdatasets package. The dataset set I have chosen to manage from the NHSRdatasets package is the NHS England accident and emergency (A&E) attendances and admissions (ae\_attendances) data. The ae\_attendances data includes reported attendances, four-hour breaches and admissions for all A&E departments in England for 2016/17 through 2018/19 (Apr-Mar). We previously selected a subset of the variables needed for my data capture tool, including period, attendances and breaches, and subsetting the data into test and training data.

```
CollectedData=read_csv(here("RawData", "CollectedDataAll.csv"))
```

```
glimpse(CollectedData)
```

### The CollectedData dataset contains:

**index:** the index column that allows us to link the data collected to the original ae\_attendances data in the 'RawData' folder.

**period:** the month that this activity relates to, stored as a date (1st of each month).

**org\_code:** the Organisation data service (ODS) code for the organisation. The ODS code is a unique code created by the Organisation data service within NHS Digital and used to identify organisations across health and social care. ODS codes are required in order to gain access to national systems like NHSmail and the Data Security and Protection Toolkit. If you want to know the organisation associated with a particular ODS code, you can look it up from the following address: <https://odspportal.digital.nhs.uk/Organisation/Search>. For example, the organisation associated with the ODS code 'AF003' is Parkway health centre.

**type:** the Department Type for this activity, either

\*1: Emergency departments are a consultant-led 24-hour service with full resuscitation facilities and designated accommodation for the reception of accident and emergency patients,

\*2: Consultant-led mono speciality accident and emergency service (e.g. ophthalmology, dental) with designated accommodation for the reception of patients, or

\*other: Other types of A&E/minor injury activity with designated accommodation for the reception of accident and emergency patients. The department may be doctor-led or nurse-led and treats at least minor injuries and illnesses and can be routinely accessed without an appointment. A service mainly or entirely

appointment-based (for example, a GP Practice or Outpatient clinic) is excluded even though it may treat a number of patients with minor illness or injury. Excludes NHS walk-in centres.(National Health Service, 2020)

**attendances:** the number of attendances for this department type at this organisation for this month.  
**breaches:** the number of attendances that breached the four-hour target.

**admissions:** the number of attendances that resulted in an admission to the hospital.(Chris Mainey, 2021)

**performance:** the performance  $([1 - \text{breaches}]/\text{attendances})$  calculated for the whole of England.

**consent:** the consent from the end-user to process and share the data collected with the data capture tool.

## Constructing a data dictionary and appending it to the data

### Build a linker data frame

#### Variable descriptions

Create a string vector representing the different variable descriptions

```
variable_description <- c("The index column that allows us to link the data collected to the original  
ae_attendances data in the 'RawData' folder.", "The month that this activity relates to, stored as a date  
(1st of each month).", "The Organisation data service (ODS) code for the organisation. If you want to know  
the organisation associated with a particular ODS code, you can look it up from the following address:  
https://odsportal.digital.nhs.uk/Organisation/Search.", "The department type for this activity.", "The  
number of attendances for this department type at this organisation for this month.", "The number of  
attendances that breached the four-hour target.",  
"The number of attendances that resulted in an admission to the hospital.", "The performance  $([1 - \text{breaches}]/\text{attendances})$  calculated for the whole of England.", "The consent from the end-user to process and  
share the data collected with the data capture tool.")  
  
print(variable_description)
```

#### Variable types

```
glimpse(CollectedData)
```

We have five quantitative values (measured values) variables and four fixed values (allowable values or codes) variables.

```
variable_type <- c(0, 1, 1, 1, 0, 0, 0, 0, 1) print(variable_type)  
linker<-build_linker(CollectedData, variable_description, variable_type)  
  
print(linker)
```

#### Data dictionary

We are now going to use the `build_dict()` function from the `dataMeta` to constructs a data dictionary for a `CollectedData` data frame with the aid of the linker data frame between.

```
dictionary <- build_dict(my.data = CollectedData, linker = linker)  
  
glimpse(dictionary)  
  
dictionary[7,4]<-“RDZ: NHS Trust - The Royal Bournemouth and Christchurch Hospitals NHS Foundation  
Trust.”  
  
dictionary[27,4] <-“other: Other types of A&E/minor injury activity with designated accommodation for the  
reception of accident and emergency patients.”
```

### Save the data dictionary for CollectedData to the ‘RawData’ folder

```
glimpse(dictionary)
write_csv(dictionary, here("RawData", "CollectedData_DataDictionary.csv"))
```

We will now incorporate attributes as metadata to the CollectedData as metadata using the ‘incorporate\_attr()’ function from the dataMeta package. For this function to run, it requires the CollectedData and dictionary and main\_string main\_string as inputs. main\_string is a character string describing the CollectedData data frame.

```
main_string <- "This data describes the NHS England accident and emergency (A&E) attendances and breaches of four-hour wait time target data from the NHSRdatasets package collected by the data capture tool." main_string
```

### Incorporate attributes as metadata

We are using the ‘incorporate\_attr()’ function to return an R dataset containing metadata stored in its attributes. The attributes we are going to add include: \* a data dictionary \* number of columns \* number of rows \* the name of the author who created the dictionary and added it, \* the time when it was last edited \* a brief description of the original dataset.

```
complete_CollectedData <- incorporate_attr(my.data = CollectedData, data.dictionary = dictionary,
main_string = main_string)
```

Change the author name

```
attributes(complete_CollectedData)$author[1]<-“B210741” complete_CollectedData attributes(complete_CollectedData)
```

### Save the CollectedData with attributes

```
save_it(complete_CollectedData, here("RawData", "complete_CollectedData"))
complete_CollectedData<-readRDS(here("RawData", "complete_CollectedData.rds"))
```

## 4. Data capture tool: Collecting data using interactive Jupyter widgets

\*Description of the code:

- Language (Python) in Jupyter
- The data is collected using interactive widgets
- The data is stored in the RawData folder