Asessment Rmd file

B209978

17/06/2022

R code to extract relevant data

link to GitHub repo

Loading packages including NHSR dataset

```
# required packages are:
library(NHSRdatasets)
library(tidyverse)
library(knitr)
library(here)
library(scales)
library(caret)
library(dataMeta)
```

Load and explore NHS England A&E attendance data

Examining structure and completeness of dataset

```
## Rows: 12,765
## Columns: 6
## $ period      <date> 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-0~
## $ org_code    <fct> RF4, RF4, RF4, R1H, R1H, R1H, AD913, RYX, RQM, RQM, RJ6, R~
## $ type        <fct> 1, 2, other, 1, 2, other, other, other, 1, other, 1, other~
## $ attendances <dbl> 21289, 813, 2850, 30210, 807, 11352, 4381, 19562, 17414, 7~
## $ breaches    <dbl> 2879, 22, 6, 5902, 11, 136, 2, 258, 2030, 86, 1322, 140, 0~
## $ admissions  <dbl> 5060, 0, 0, 6943, 0, 0, 0, 0, 3597, 0, 2202, 0, 0, 0, 3360~
```

```
## # A tibble: 12,765 x 6
##    period     org_code type  attendances breaches admissions
##    <date>     <fct>    <fct>       <dbl>    <dbl>      <dbl>
##  1 2017-03-01 RF4      1           21289     2879       5060
##  2 2017-03-01 RF4      2             813       22          0
##  3 2017-03-01 RF4      other        2850        6          0
##  4 2017-03-01 R1H      1           30210     5902       6943
##  5 2017-03-01 R1H      2             807       11          0
##  6 2017-03-01 R1H      other       11352      136          0
##  7 2017-03-01 AD913    other        4381        2          0
##  8 2017-03-01 RYX      other       19562      258          0
##  9 2017-03-01 RQM      1           17414     2030       3597
## 10 2017-03-01 RQM      other        7817       86          0
## # ... with 12,755 more rows
```

```
## $period
## [1] 0
##
## $org_code
## [1] 0
##
## $type
## [1] 0
##
## $attendances
## [1] 0
##
## $breaches
## [1] 0
##
## $admissions
## [1] 0
```

Adding index for later linkage

index	period	org_code	type	attendances	breaches	admissions
1	Mar-17	RF4	1	21,289.0	2,879.0	5,060.0
2	Mar-17	RF4	2	813.0	22.0	0.0
3	Mar-17	RF4	other	2,850.0	6.0	0.0
4	Mar-17	R1H	1	30,210.0	5,902.0	6,943.0
5	Mar-17	R1H	2	807.0	11.0	0.0
6	Mar-17	R1H	other	11,352.0	136.0	0.0
7	Mar-17	AD913	other	4,381.0	2.0	0.0
8	Mar-17	RYX	other	19,562.0	258.0	0.0
9	Mar-17	RQM	1	17,414.0	2,030.0	3,597.0
10	Mar-17	RQM	other	7,817.0	86.0	0.0

Filter data

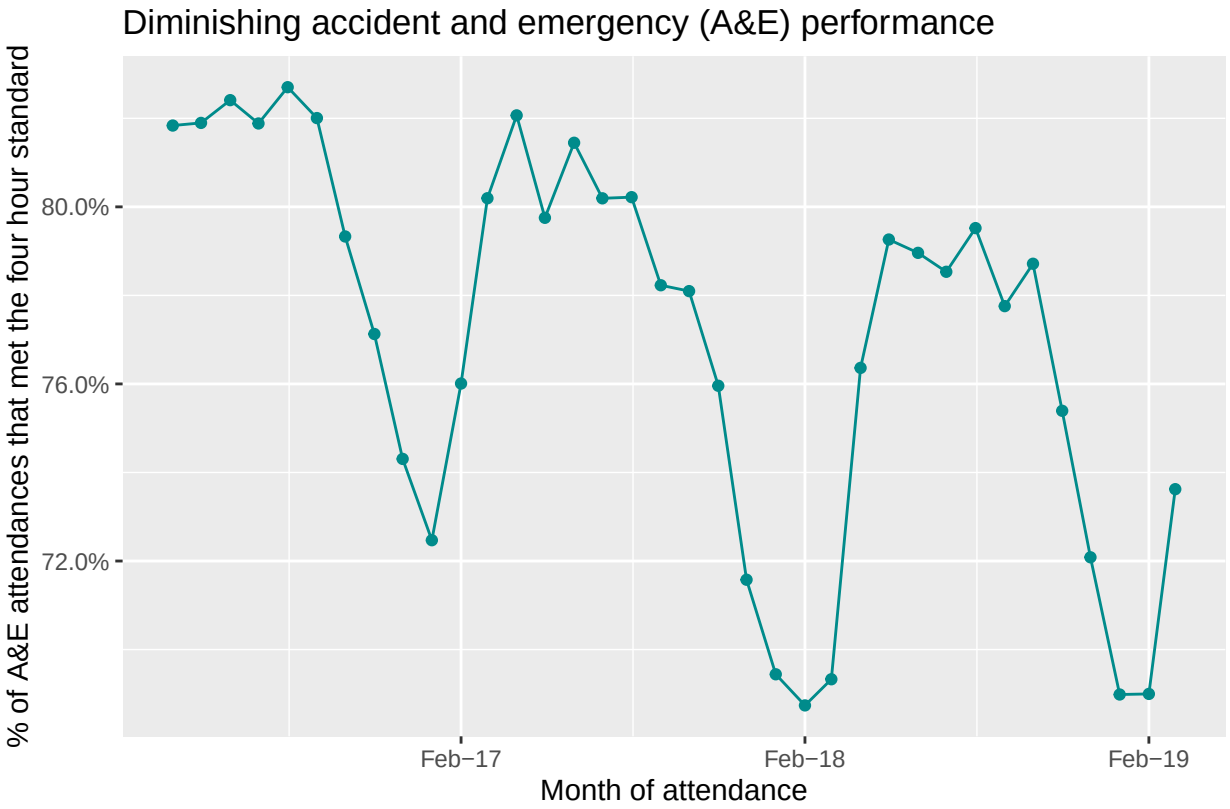
Removed all hospitals except consultant led Emergency Departments covered by selected ambulance service

```
## # A tibble: 468 x 6
##    index period     org_code attendances breaches admissions
##    <int> <date>     <fct>          <dbl>    <dbl>      <dbl>
##  1    77 2017-03-01 RLT             6726      799       1119
##  2   117 2017-03-01 RXK            14665     2828       3067
##  3   122 2017-03-01 RXW            10371     2217       2441
##  4   126 2017-03-01 RJC             5572      198       1806
##  5   131 2017-03-01 RNA             9006      679       2425
##  6   133 2017-03-01 RQW             8737     1935       1835
##  7   135 2017-03-01 RL4            11505     1651       2606
##  8   138 2017-03-01 RRK            10032     1543       2818
##  9   139 2017-03-01 RKB            11986     3213       4077
## 10   145 2017-03-01 RJE            15129     4019       4511
## # ... with 458 more rows
```

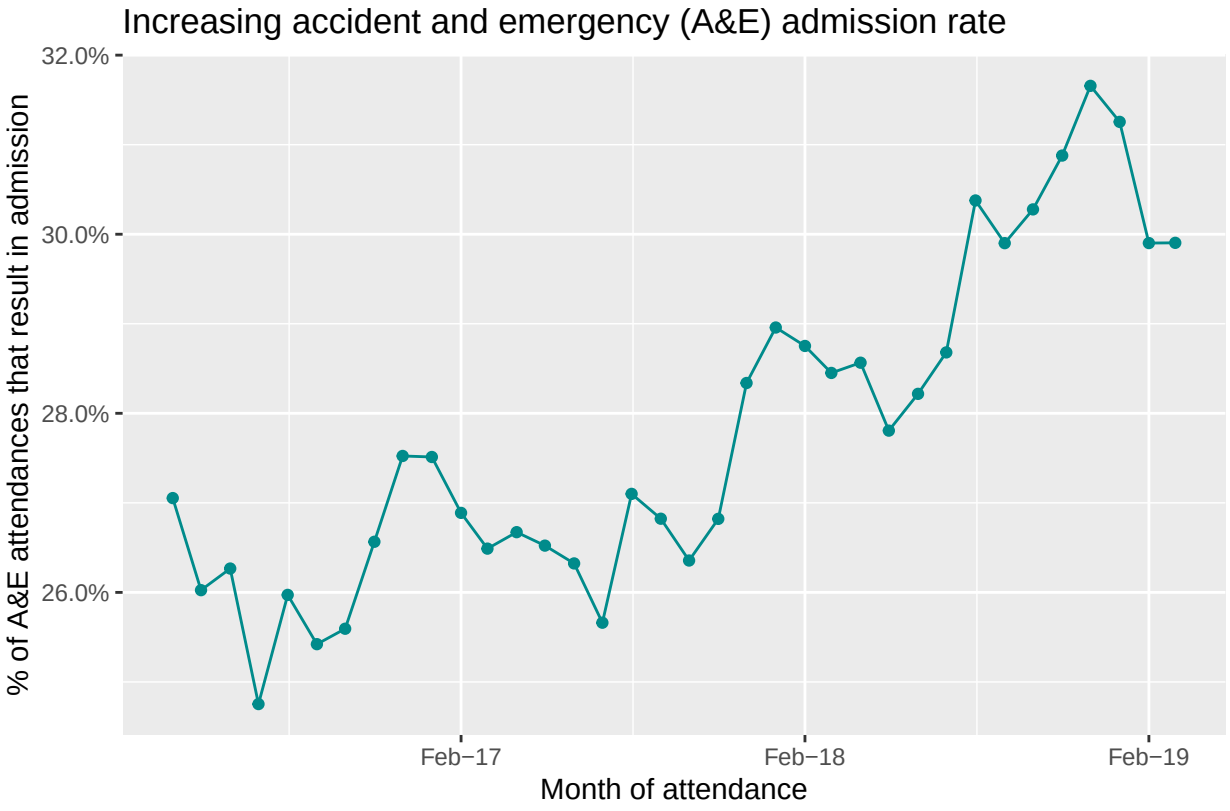
Calculate metrics

```
## Rows: 36
## Columns: 6
## $ period             <date> 2016-04-01, 2016-05-01, 2016-06-01, 2016-07-01, 20~
## $ attendances        <dbl> 115395, 127489, 120287, 127278, 117611, 119048, 124~
## $ breaches           <dbl> 20959, 23080, 21158, 23058, 20343, 21422, 25647, 27~
## $ admissions         <dbl> 31218, 33179, 31595, 31505, 30546, 30264, 31758, 31~
## $ breach_performance <dbl> 0.8183717, 0.8189648, 0.8241040, 0.8188375, 0.82703~
## $ admission_rate     <dbl> 0.2705317, 0.2602499, 0.2626635, 0.2475290, 0.25972~
```

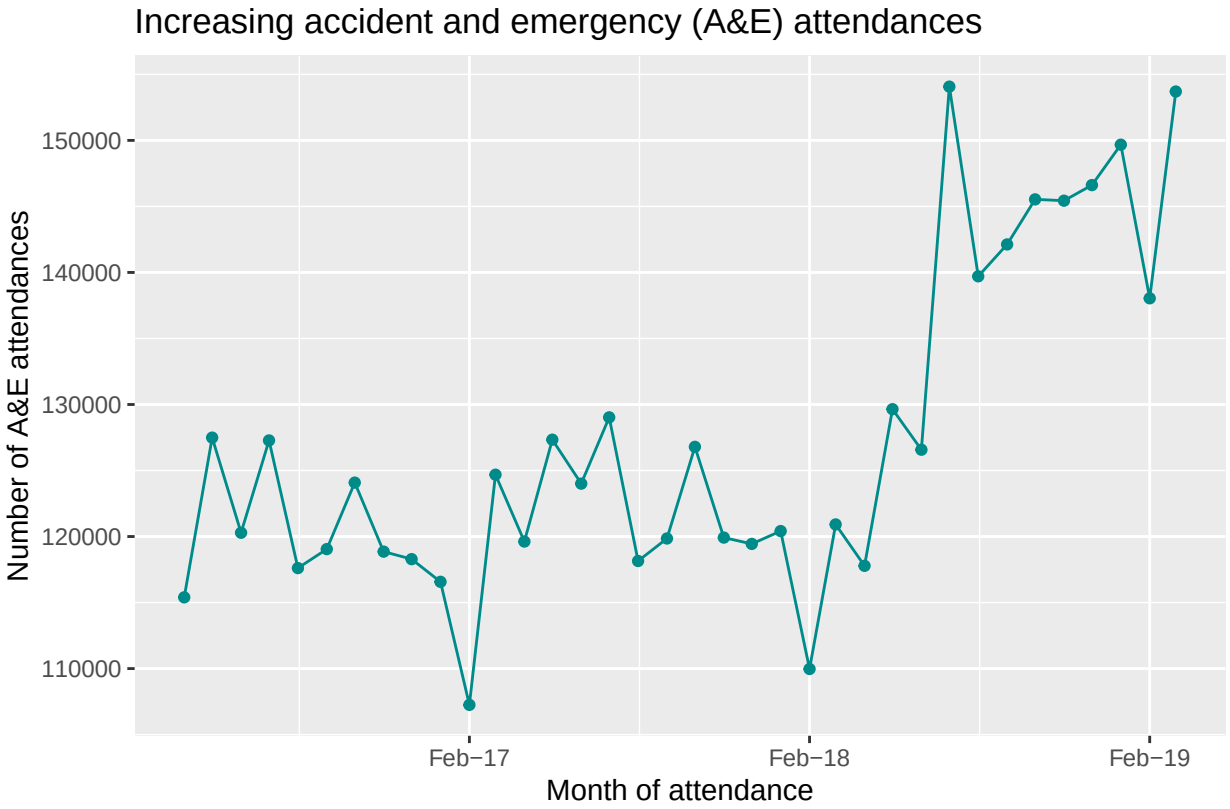
Brief visualisation of regional patterns



Source: NHSRdatasets



Source: NHSRdatasets



Save data subset

```
## Rows: 468
## Columns: 6
## $ index       <int> 77, 117, 122, 126, 131, 133, 135, 138, 139, 145, 148, 154,~
## $ period      <date> 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-01, 2017-03-0~
## $ org_code    <fct> RLT, RXK, RXW, RJC, RNA, RQW, RL4, RRK, RKB, RJE, RBK, RWP~
## $ attendances <dbl> 6726, 14665, 10371, 5572, 9006, 8737, 11505, 10032, 11986,~
## $ breaches    <dbl> 799, 2828, 2217, 198, 679, 1935, 1651, 1543, 3213, 4019, 1~
## $ admissions  <dbl> 1119, 3067, 2441, 1806, 2425, 1835, 2606, 2818, 4077, 4511~
```

Divide dataset into training, marker and test

index	period	org_code	attendances	breaches	admissions
77	Mar-17	RLT	6,726.0	799.0	1119
117	Mar-17	RXK	14,665.0	2,828.0	3067
122	Mar-17	RXW	10,371.0	2,217.0	2441
126	Mar-17	RJC	5,572.0	198.0	1806
131	Mar-17	RNA	9,006.0	679.0	2425
133	Mar-17	RQW	8,737.0	1,935.0	1835
135	Mar-17	RL4	11,505.0	1,651.0	2606
138	Mar-17	RRK	10,032.0	1,543.0	2818
139	Mar-17	RKB	11,986.0	3,213.0	4077
145	Mar-17	RJE	15,129.0	4,019.0	4511

index	period	org_code	attendances	breaches	admissions
2550	Aug-16	RNA	8,575	603	2657

index	period	org_code	attendances	breaches	admissions
2881	Jul-16	RXK	14,488	2,128	3141
2896	Jul-16	RNA	8,947	596	2599
4258	Mar-18	RXK	13,805	3,556	3429
4281	Mar-18	RRK	9,936	2,154	3896
5043	Jan-18	RLQ	4,532	1,263	1437
6471	Sep-17	RWP	9,817	2,716	2921
7137	Jul-17	RJC	5,811	297	1617
7509	Jun-17	RWP	10,313	2,824	3174
9577	Dec-18	RXK	13,604	4,432	3744
10327	Oct-18	RKB	12,519	1,937	4407

Data dictionary

Read in data collected in Python

```
## Rows: 11
## Columns: 9
## $ index              <dbl> 2881, 2896, 4258, 4281, 5043, 6471, 7137, 7509, 957~
## $ period             <date> 2016-07-01, 2016-07-01, 2018-03-01, 2018-03-01, 20~
## $ org_code           <chr> "RXK", "RNA", "RXK", "RRK", "RLQ", "RWP", "RJC", "R~
## $ attendances        <dbl> 1488, 8947, 13805, 9936, 4532, 9817, 5811, 10313, 1~
## $ breaches           <dbl> 2128, 596, 3556, 2154, 1263, 2716, 297, 2824, 4432,~
## $ admissions         <dbl> 3141, 2599, 3429, 3896, 1437, 2921, 1617, 3174, 374~
## $ breach_performance <dbl> -0.4301075, 0.9333855, 0.7424122, 0.7832126, 0.3170~
## $ admission_rate     <dbl> 2.1108871, 0.2904884, 0.2483883, 0.3921095, 0.31707~
## $ consent            <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRU~
```

Build linker data frame with variable descriptions and types

```
## [1] "The index column that allows us to link the data collected to the original ae_attendances data i
## [2] "The month that this activity relates to, stored as a date (1st of each month)."
## [3] "The Organisation data service (ODS) code for the organisation. If you want to know the organisat
## [4] "The number of attendances for this department type at this organisation for this month."
## [5] "The number of attendances that breached the four-hour target."
## [6] "The number of attendances that resulted in an admission to the hospital."
## [7] "The performance ([1 - breaches]/attendances)"
## [8] "The rate of admission (admissions/attendances)"
## [9] "The consent from the end-user to process and share the data collected with the data capture tool
```

```
## [1] 0 1 1 0 0 0 0 0 1
```

```
##             var_name
## 1              index
## 2             period
## 3           org_code
## 4         attendances
## 5           breaches
## 6          admissions
```

```
## 7 breach_performance
## 8     admission_rate
## 9            consent
##
## 1
## 2
## 3 The Organisation data service (ODS) code for the organisation. If you want to know the organisatio
## 4
## 5
## 6
## 7
## 8
## 9
##   var_type
## 1        0
## 2        1
## 3        1
## 4        0
## 5        0
## 6        0
## 7        0
## 8        0
## 9        1
```

Use linker dataframe to create data dictionary

```
##        variable name
## 1     admission_rate
## 2         admissions
## 3        attendances
## 4 breach_performance
## 5           breaches
## 6            consent
##                                                                                 variable description
## 1                                                    The rate of admission (admissions/attendances)
## 2                           The number of attendances that resulted in an admission to the hospital.
## 3          The number of attendances for this department type at this organisation for this month.
## 4                                                      The performance ([1 - breaches]/attendances)
## 5                                    The number of attendances that breached the four-hour target.
## 6 The consent from the end-user to process and share the data collected with the data capture tool.
##                          variable options notes
## 1  0.237557194882809 to 2.11088709677419
## 2                           1437 to 4407
## 3                          1488 to 13805
## 4 -0.43010752688172 to 0.933385492343802
## 5                            297 to 4432
## 6                                   TRUE
```

Appending data dictionary to collected data

```
## [1] "This data describes accident and emergency (A&E) metrics for hospital trusts within the chosen a
```

```
## # A tibble: 11 x 9
##    index period     org_code attendances breaches admissions breach_performance
##  * <dbl> <date>     <chr>          <dbl>    <dbl>      <dbl>              <dbl>
```

```
##  1  2881 2016-07-01 RXK             1488     2128       3141             -0.430
##  2  2896 2016-07-01 RNA             8947      596       2599              0.933
##  3  4258 2018-03-01 RXK            13805     3556       3429              0.742
##  4  4281 2018-03-01 RRK             9936     2154       3896              0.783
##  5  5043 2018-01-01 RLQ             4532     1263       1437              0.317
##  6  6471 2017-09-01 RWP             9817     2716       2921              0.298
##  7  7137 2017-07-01 RJC             5811      297       1617              0.278
##  8  7509 2017-06-01 RWP            10313     2824       3174              0.308
##  9  9577 2018-12-01 RXK            13604     4432       3744              0.275
## 10 10327 2018-10-01 RKB            12519     1937       4407              0.352
## 11 12530 2018-04-01 RL4            10709     1704       2544              0.238
## # ... with 2 more variables: admission_rate <dbl>, consent <lgl>
```

```
## $row.names
##  [1]  1  2  3  4  5  6  7  8  9 10 11
##
## $names
## [1] "index"              "period"             "org_code"
## [4] "attendances"        "breaches"           "admissions"
## [7] "breach_performance" "admission_rate"     "consent"
##
## $spec
## cols(
##   index = col_double(),
##   period = col_date(format = ""),
##   org_code = col_character(),
##   attendances = col_double(),
##   breaches = col_double(),
##   admissions = col_double(),
##   breach_performance = col_double(),
##   admission_rate = col_double(),
##   consent = col_logical()
## )
##
## $problems
## <pointer: 0x564f4c770810>
##
## $class
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
##
## $main
## [1] "This data describes accident and emergency (A&E) metrics for hospital trusts within the chosen a
##
## $dictionary
##         variable name
## 1      admission_rate
## 2          admissions
## 3         attendances
## 4  breach_performance
## 5            breaches
## 6             consent
## 7               index
## 8            org_code
## 9
```

```
## 10
## 11
## 12
## 13
## 14
## 15
## 16             period
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8  The Organisation data service (ODS) code for the organisation. If you want to know the organisatio
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
##                           variable options notes
## 1   0.237557194882809 to 2.11088709677419
## 2                          1437 to 4407
## 3                         1488 to 13805
## 4  -0.43010752688172 to 0.933385492343802
## 5                           297 to 4432
## 6                                  TRUE
## 7                         2881 to 12530
## 8                                   RXK
## 9                                   RNA
## 10                                  RRK
## 11                                  RLQ
## 12                                  RWP
## 13                                  RJC
```

```
## 14                                    RKB
## 15                                    RL4
## 16                                  16983
## 17                                  17591
## 18                                  17532
## 19                                  17410
## 20                                  17348
## 21                                  17318
## 22                                  17866
## 23                                  17805
## 24                                  17622
##
## $last_edit_date
## [1] "2022-06-24 15:48:26 UTC"
##
## $author
## [1] "B209978"
```

Data capture tool (Python)

Perfomed in Python. Jupyter widgets designed to capture data from relevant hospitals. Performance and admission rate metrics calculated and inputted using simple block of Python code.