We identified three objectives: Identify trends in the data, Identify ICBs where NHS is the busiest and where it is most inefficient, Provide a set of tools for the users of the analysis to do bespoke analysis. We use average delay between booking and appointment, as a measure of strain in an ICB. We complement the dataset provided by the client with additional data publicly available on NHS's patients and staff (FTE) data.We would recommend that inefficient ICBs shorten their appointment to handle a higher load, and increase staff per patient some in underperforming ICBs and in places with the oldest population. Lastly, we recommend that we ask the management of Derby & Derbyshire; Bedfordshire, Luton & Milton Keynes; Nottingham & Nottinghamshire; Dorset ICBs for an explanation of poor performance.

Analysis of NHS Appointment Data (January 2020 and June 2022)

PB Analytics Ltd. (Patryk Basiewicz)

Table of Contents

Background/Context of the Business Scenario

We were tasked with analysing NHS data on appointments between January 2020 and June 2022. Hence, our period of analysis covers the COVID pandemic. We identified three objectives:

1. Identifying trends in the data, using different properties of the appointments
2. Identify ICBs where NHS is the busiest and where it is most inefficient
3. Provide a set of tools for the users of the analysis to do bespoke analysis.

We complement the dataset provided by the client with additional data publicly available on NHS's patients and staff (FTE) data.

Analytical approach

Additional Data

I augment the given data with additional datasets. They are shown below. One file is a key for all the regions, three of the files are GeoJSONs for producing maps, and one has information on patients and NHS FTEs for each region. I do note that patient/FTE data is more recent, but I assume that proportions have not changed too much.

Figure 1 Additional Datasets

Dataset	Description	Mode of Access
Sub ICB Locations to Integrated Care Boards to NHS England (Region) (July 2022) Lookup in England	A key that has all the Regions, ICBs, and Sub-ICB data, and the associated codes	API
NHS England (Regions) (July 2022) EN BFC	A data set GeoJSON file of NHS Regions to be used with Geopandas	API
Integrated Care Boards (April 2023) EN BSC	A data set GeoJSON file of NHS ICBs to be used with Geopandas	API
Sub Integrated Care Board Locations (April 2023) EN BFC	A data set GeoJSON file of NHS Sub-ICBs to be used with Geopandas	API
General Practice Workforce 30 September 2023	An excel dashboard that contains breakdown of patients (by age) and NHS FTEs by sub-ICB	Online file

Key variables

We initially use an average number of monthly appointments as a measure of business, but in the deeper analysis, we scale it by the number of FTEs or patients. To get better measures of business, we compute average delay (over a month, place, or both) using the time_between_book_and_appointment category (in AR) and average duration (over a month, place, or both) using the actual_duration category (in AD[1]), using the map shown below. The idea is that places that have longer duration before the appointment and have shorter appointments are more busy.

Figure 2 Mapping of Average Delay and Average Delay Variables

Mapping for Average Delay	
Category	**Days**
Same Day	0
1 Day	1
2 to 7 Days	4.5
8 to 14 Days	11
15 to 21 Days	18
1 Day	1
More than 28 Days	35
Unknown / Data Quality	Nan

Mapping for Average Duration	
Category	**Minutes**
1-5 Minutes	3
6-10 Minutes	8
11-15 Minutes	13
16-20 Minutes	19
21-30 Minutes	25
6-10 Minutes	8
Unknown / Data Quality	'Nan'

We start by importing the three datasets into corresponding Dataframes. Because our main variable is an average delay, AR does not have sub-ICB granularity; we focus mostly on analysis at the ICB level and above.

We format so they all have the same information: all coded and names for (Sub) ICBs and Regions in the file (if possible), and have monthly time series variable. I also reduce the size of some categories by grouping them together. See the Appendix.

The initial set of results uses just these files. The next set of merges joins the three primary files: data on patients and data on NHS FTES on ICBs. That means I have to pivot the Staff file to get different types of results.

From the primary data file, I only keep the two variables above. After merging, We calculate variables such as appointments per month per patient or patients over 65 years old per GP. There are many others.

Because we lose some of the data on the type of appointments, we also join NC[2] and AR on months and ICBs, but I keep all the additional variables (such as Context Type & Context Type). To do that, I have to pivot the DataFrames and make mergers with multiple levels of columns and indices.

[1] AD stands for actual_duration.csv
[2] NC stand for national_category.csv

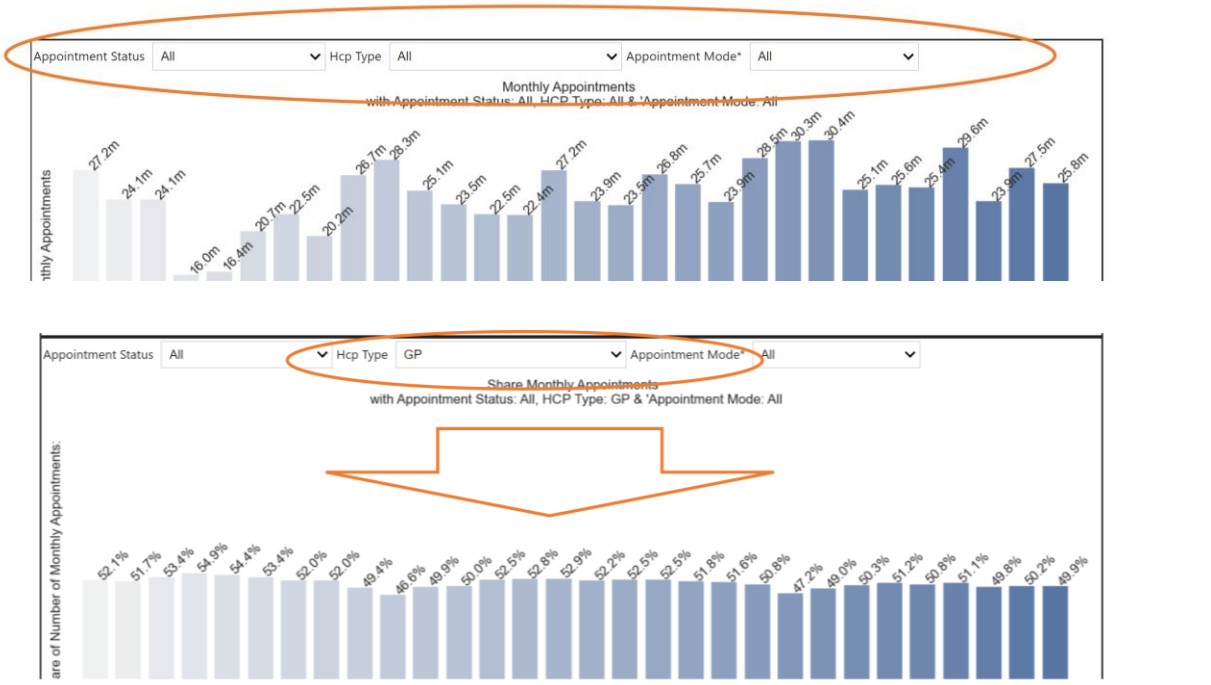
Most importantly, since we pivoted with this assignment, we presented all the charts and visualisation in an interactive manner. In order to capture iterations, we used ipywidgets. We use geopandas for map visualisations. (We will opt for another standard in future, as geopandas is slow.

Lastly, we analysed Twitter data but found nothing useful.

Visualisations and Insights

The analysis is split into three sets of visualisations and tools. In the first one, we provide tools that allow the user to observe the time series evolution of variables of interest. They mainly consist of a chart shown below that allows the user to filter concurrently with up to three categories. The tools based on the count of appointments also have a version that shows the percentage of each category each month.s

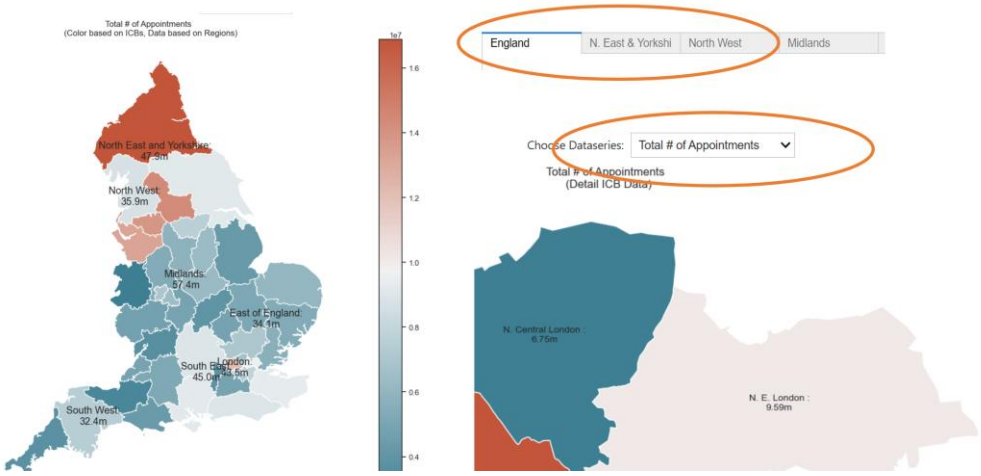
Figure 3 Example of Time Series Charts



The second set of visualisations analyses the cross-sectional variables obtained from a merge of the three primary datasets, the patient data & NHS FTE data, as well as calculated variables from the variables in those datasets. There are nine tabs in the figure. Each has a map. The first one is the map for England using ICB level data (but with Regional labels). The next seven each contain a region with the associated data. The last one shows England again, but using Sub-ICB data[3]. We split it out like that as it is easier to read and faster to compute. Remember, there are about 30 variables that the user can toggle to see how the given metric compares across England.

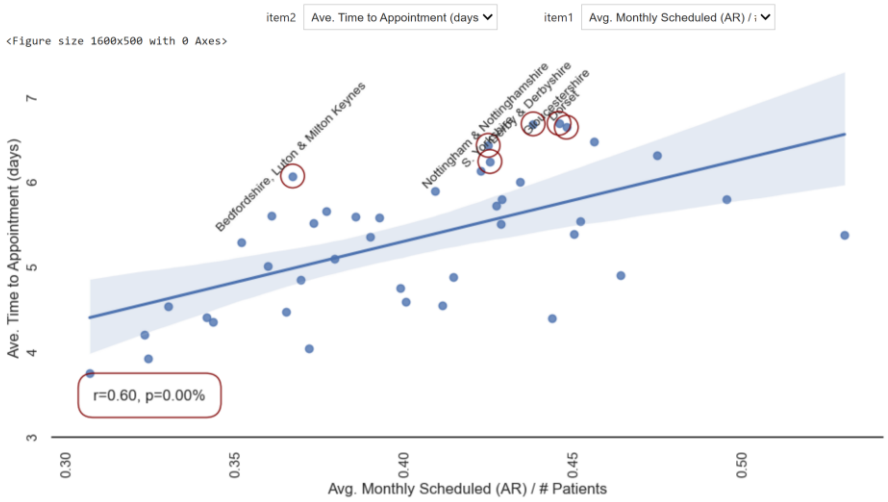
[3] This set does not include data from AR, so it does not have the key variable of average delay.

Figure 4 Set of Screenshots and UI of the Geographic Analysis



The third analysis is a bivariate extension of the analysis above. The user can choose two variables from the cross-sectional data for different ICBs and analyse the relationship between them. The chart also shows the correlation between those two variables and the p-value associated with this relationship.

Figure 5 Example of a Bivariate Analysis with Switches



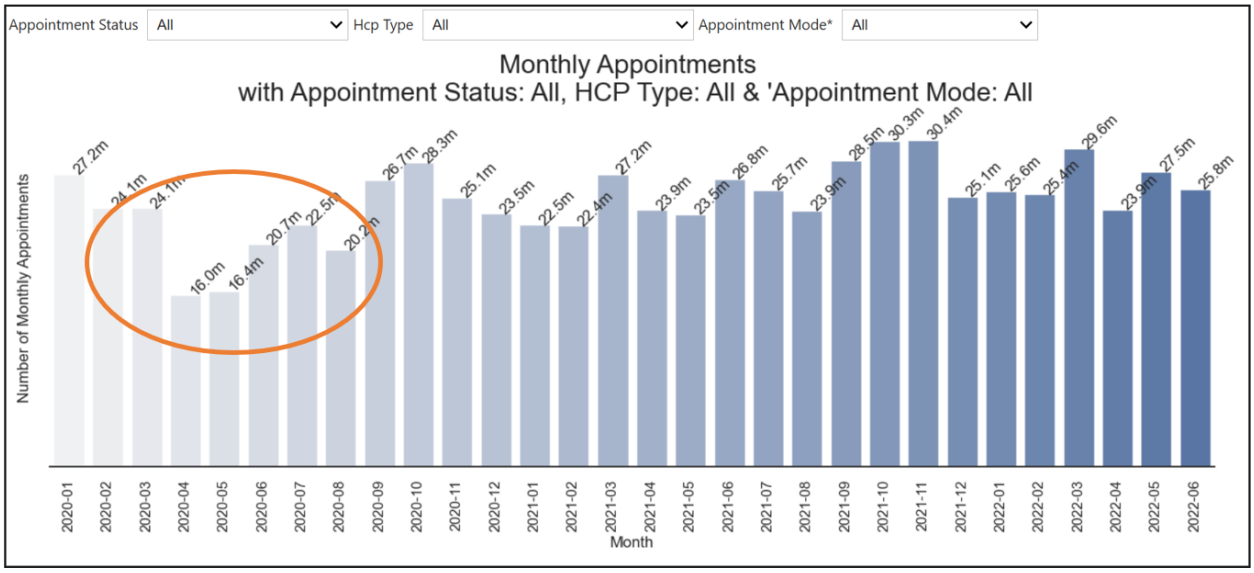
Patterns and predictions

Time Series Patterns

Given that we built a set of tools for analysing the data, we present only a subset of insights that we found interesting, and we suggest that the client use those tools to find other insights that fit their changing business needs.

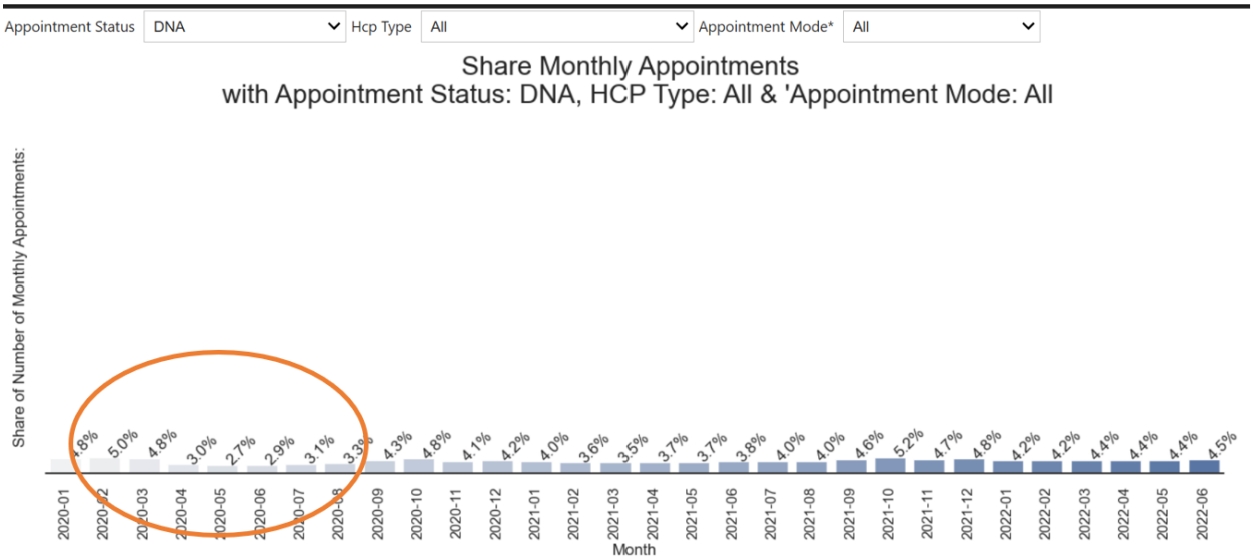
First, we see a significant fall in appointments during the COVID outbreak.

Figure 6 Evolution of Number of Appointments



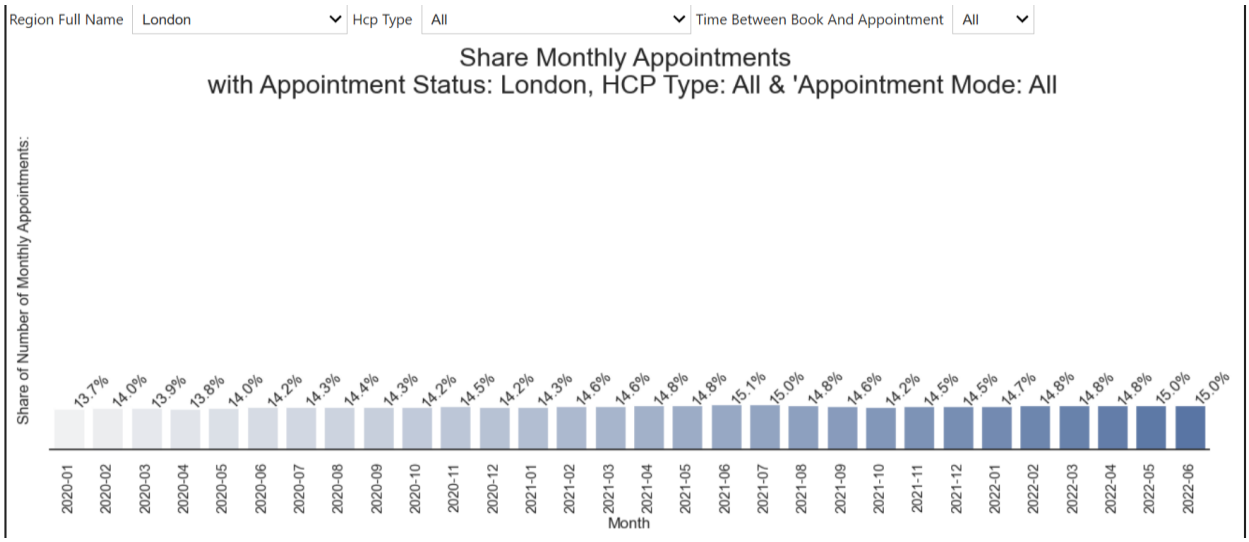
However, the incidence of missed appointments also fell during COVID-19 and was flat for GP appointments (not shown.)

Figure 7 Evolution of Monthly Appointment (Not Attended Only)



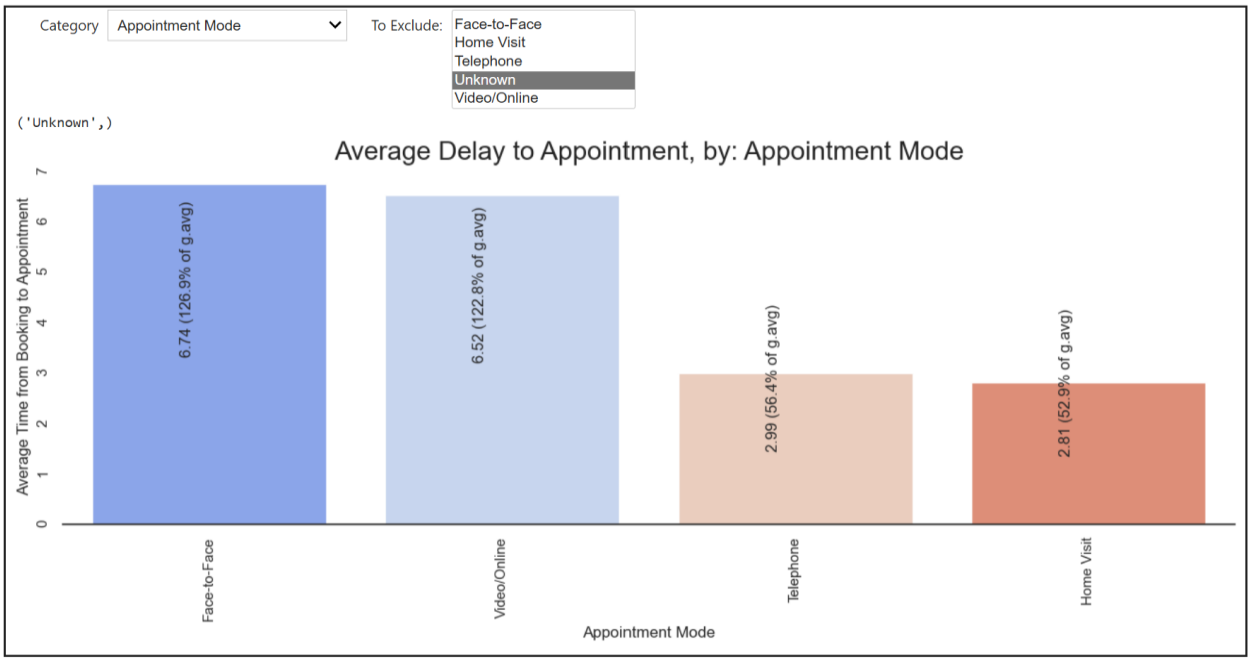
In terms of regional shift, London stands out as the only change.

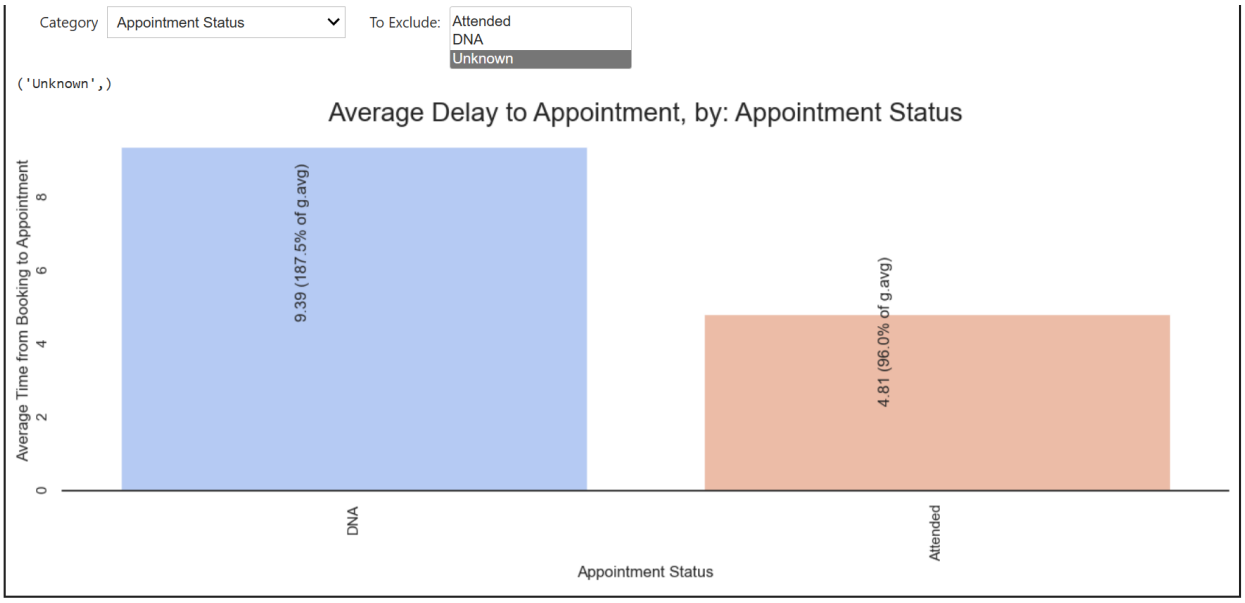
Figure 8 Evolution of Number of Appointments (London Only)

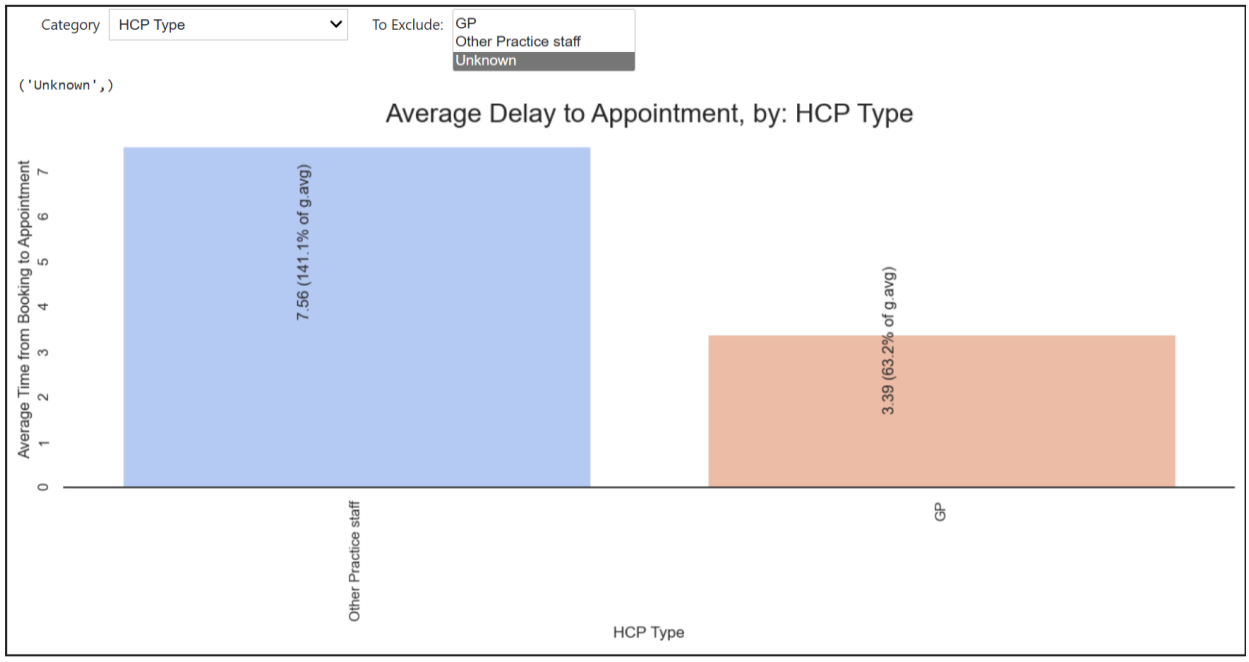


Secondly, our measure of service strain differs by Appointment Mode and type of Health Care Professional. It is much larger for non-GP appointments and those that are not attended.

Figure 9 Average Time between Booking and Appointment

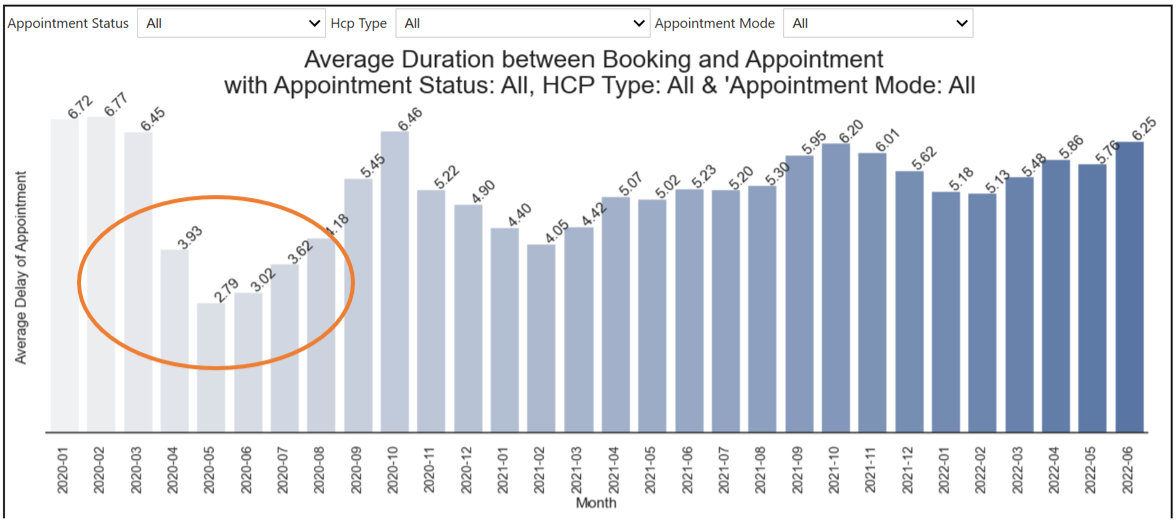


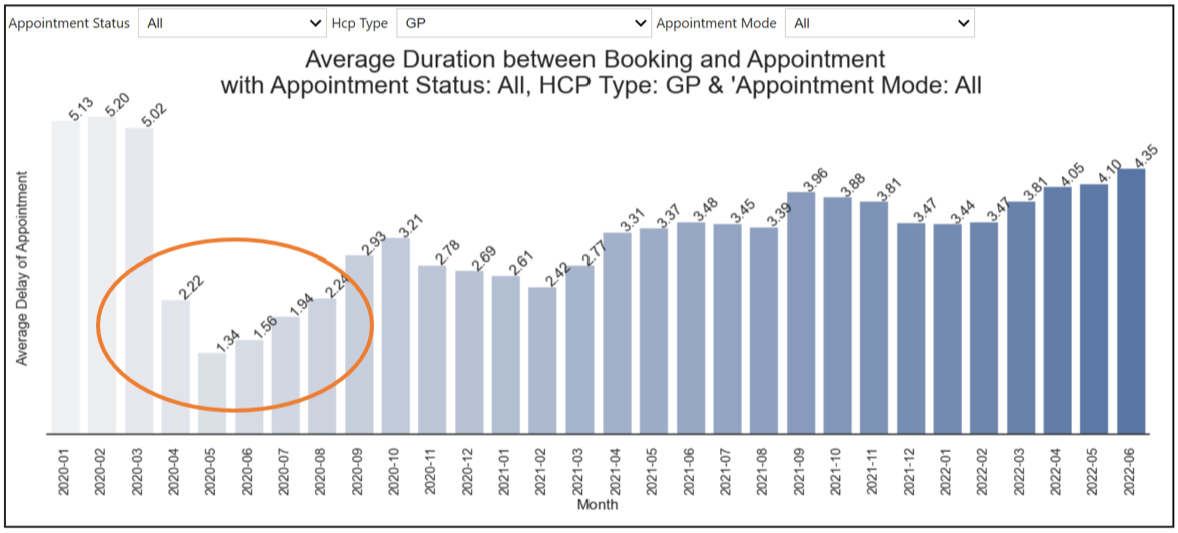


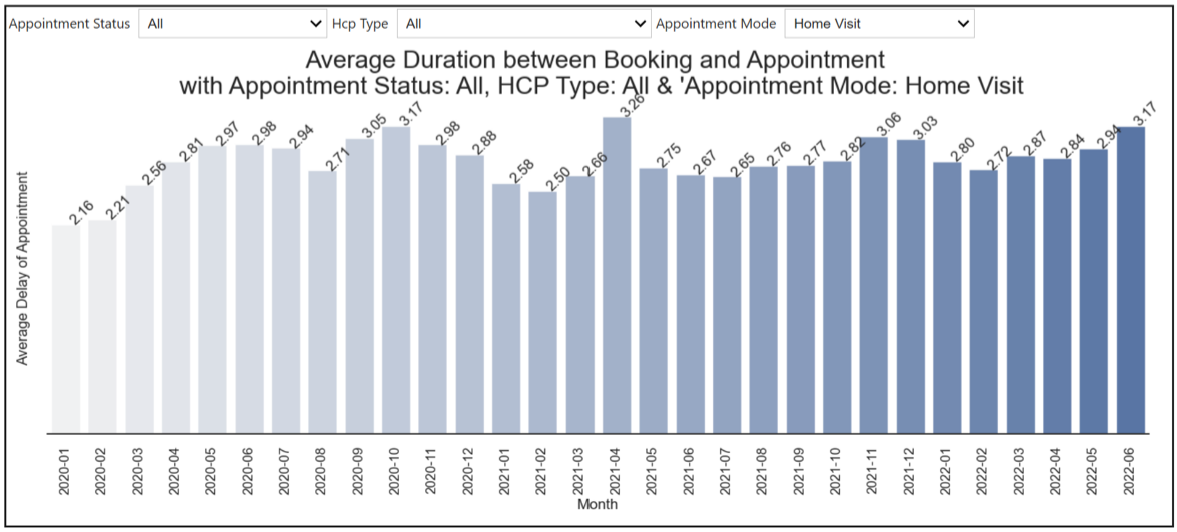


The time series shows that the average delay fell during the COVID pandemic, especially for GP Appointments, but was unchanged for Home Visits.

Figure 10 Evolution of Time between Booking and Appointment (see Switches for Variables)



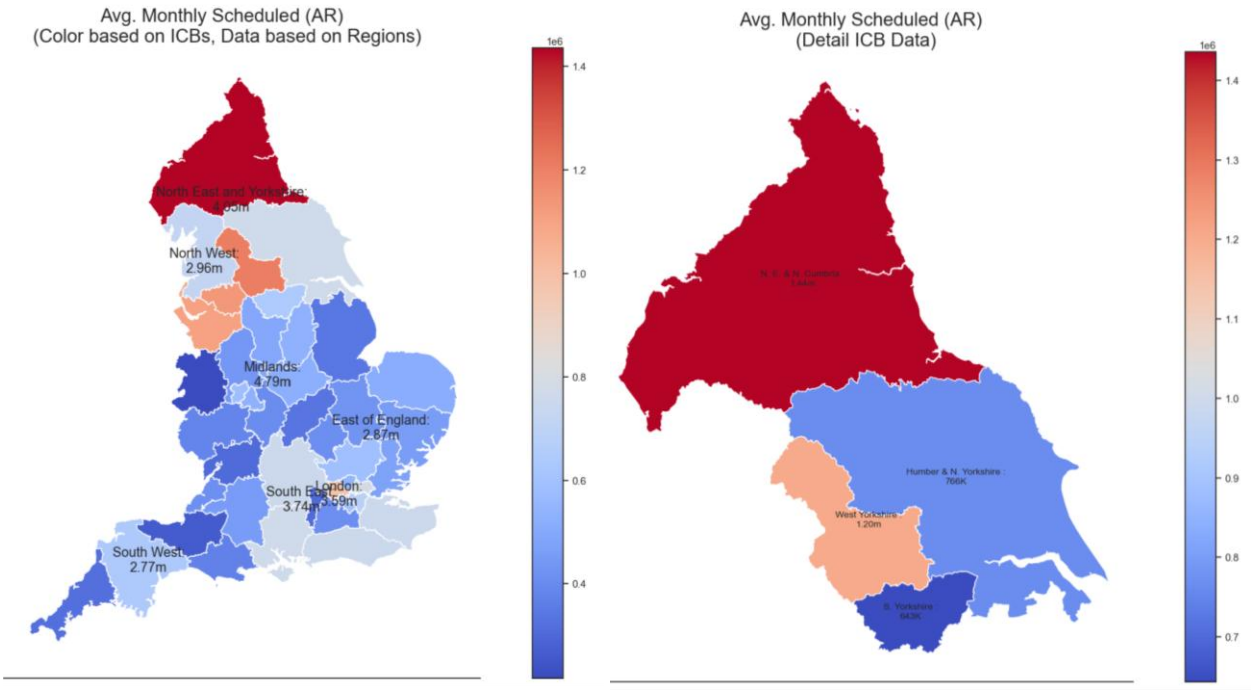




Regional Analysis Patterns

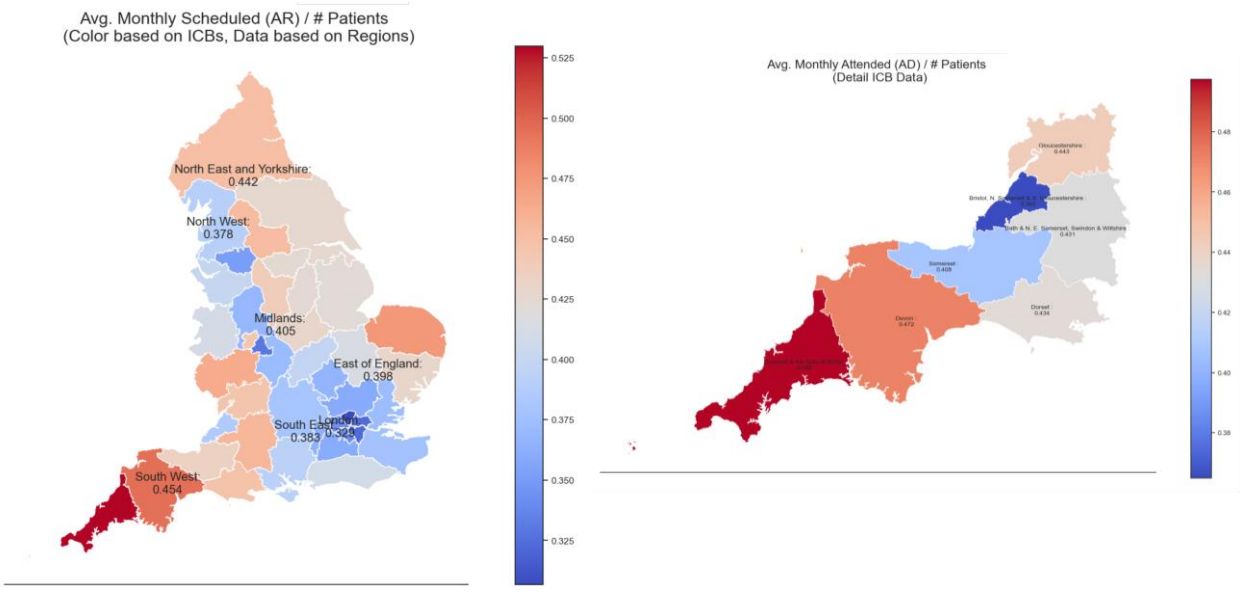
North East & Yorksire and North East seem to be the most busy with Cumbria, Manchester and Cheshire are most busy.

Figure 11 Ave. Monthly Appointments



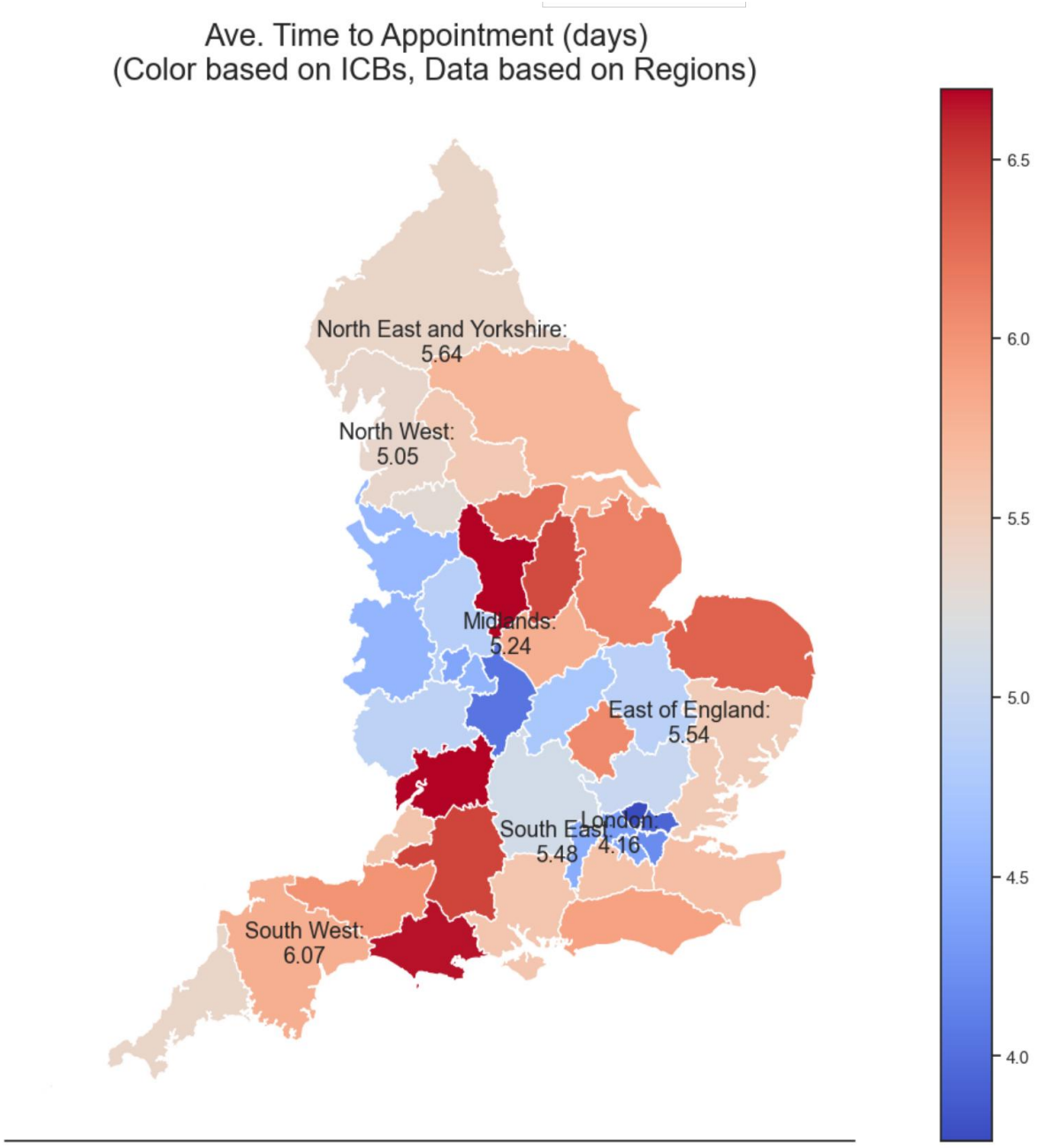
But, the picture changes when we scale by patients, with eastern ICBs of the West performing very poorly.

Figure 12 Ave. Monthly Appointments per Patient



However, as measured by average delay, strain seems to be spread out across England, with London and parts of western England doing well.

Figure 13 Average Time between Booking and Appointment

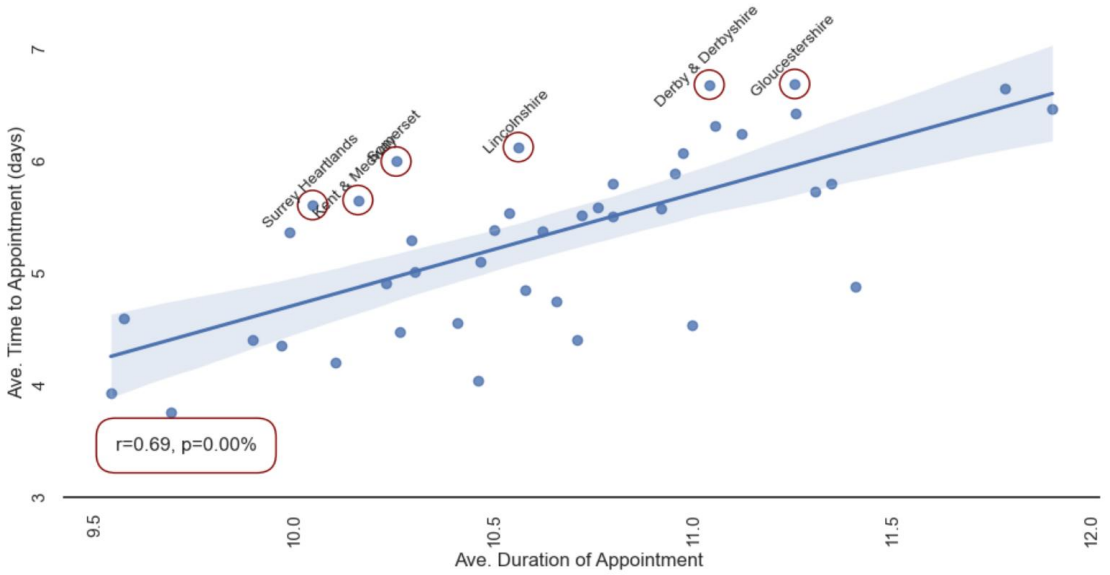


Interestingly, as shown in the appendix, London has the most Patients per NHS FTE in England but scores among the worst for Patients (over 65) per GP (i.e. the most sickly patients and most effective staff)

Recommendations based on Bivariate Analysis

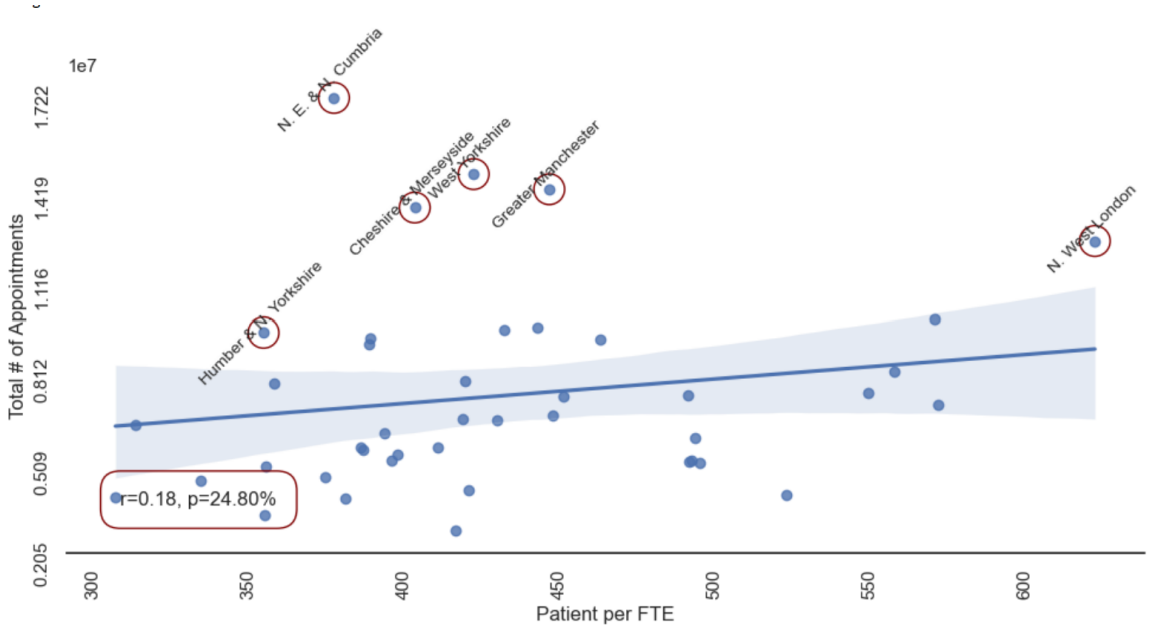
Firstly, it seems that longer waits for appointments are related to longer appointments, **so we would recommend that inefficient ICBs shorten their appointment to handle higher loads.**

Figure 14



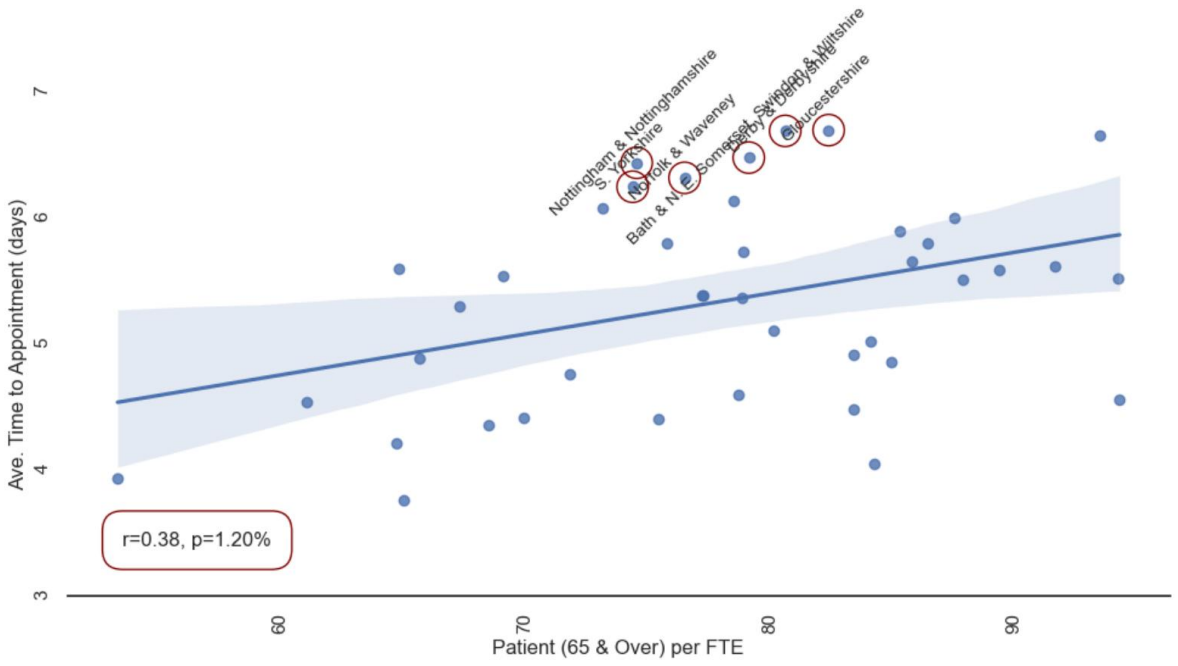
Second, ICBs' amount of appointments does not correlate to patients per FTE, and it seems that FTE shortage is related to higher delays to appointments, so we would recommend increasing staff per patient for **some underperforming ICBs.**

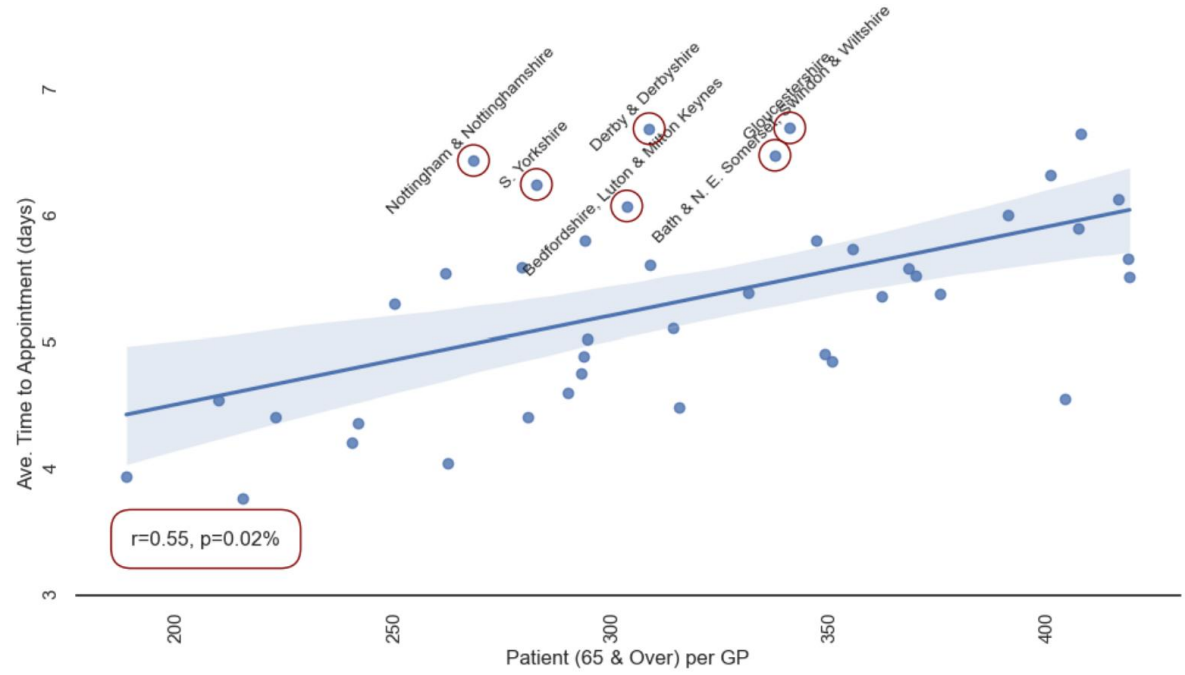
Figure 15 Total Appointments vs Patient per FTE



Thirdly, time to appointment is strongly related to the age of the patients in the ICB, **so we would recommend increasing staff in places with the oldest population.**

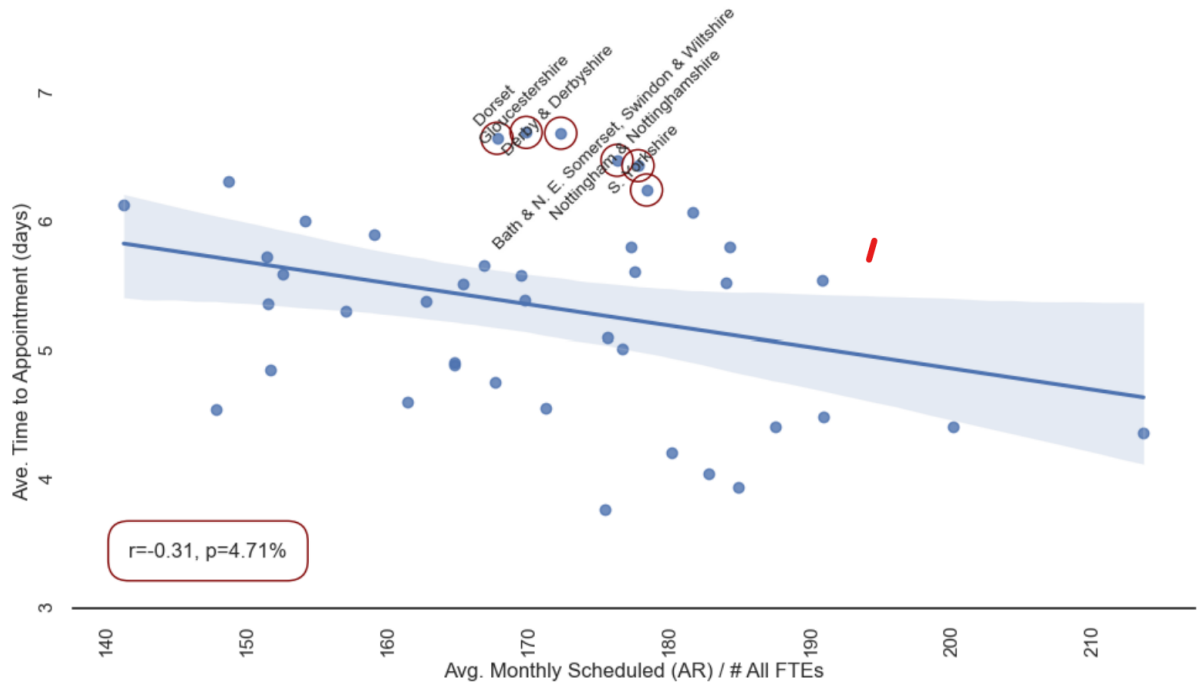
Figure 16 Average Delay to Appointment vs Age of Patients





Lastly, there are a number of ICBs that underperform consistently. Those are Derby & Derbyshire; Bedfordshire, Luton & Milton Keynes; Nottingham & Nottinghamshire; Dorset; Gloucestershire; & S. Yorkshire. **We would recommend that we ask the management of those ICBs for an explanation.**

Figure 17 Total Appointments vs Measure of Busyness



APPENDIX

Below is a mapping of Categories in some variables that are very small made the analysis noisy.

Figure 18 Mapping of Categories the Are Small

Old Category	New Category
Face-to-Face	Face-to-Face *
Home Visit	Face-to-Face *
Telephone	Telephone*
Video/Online	Telephone*
Unknown	Unknown

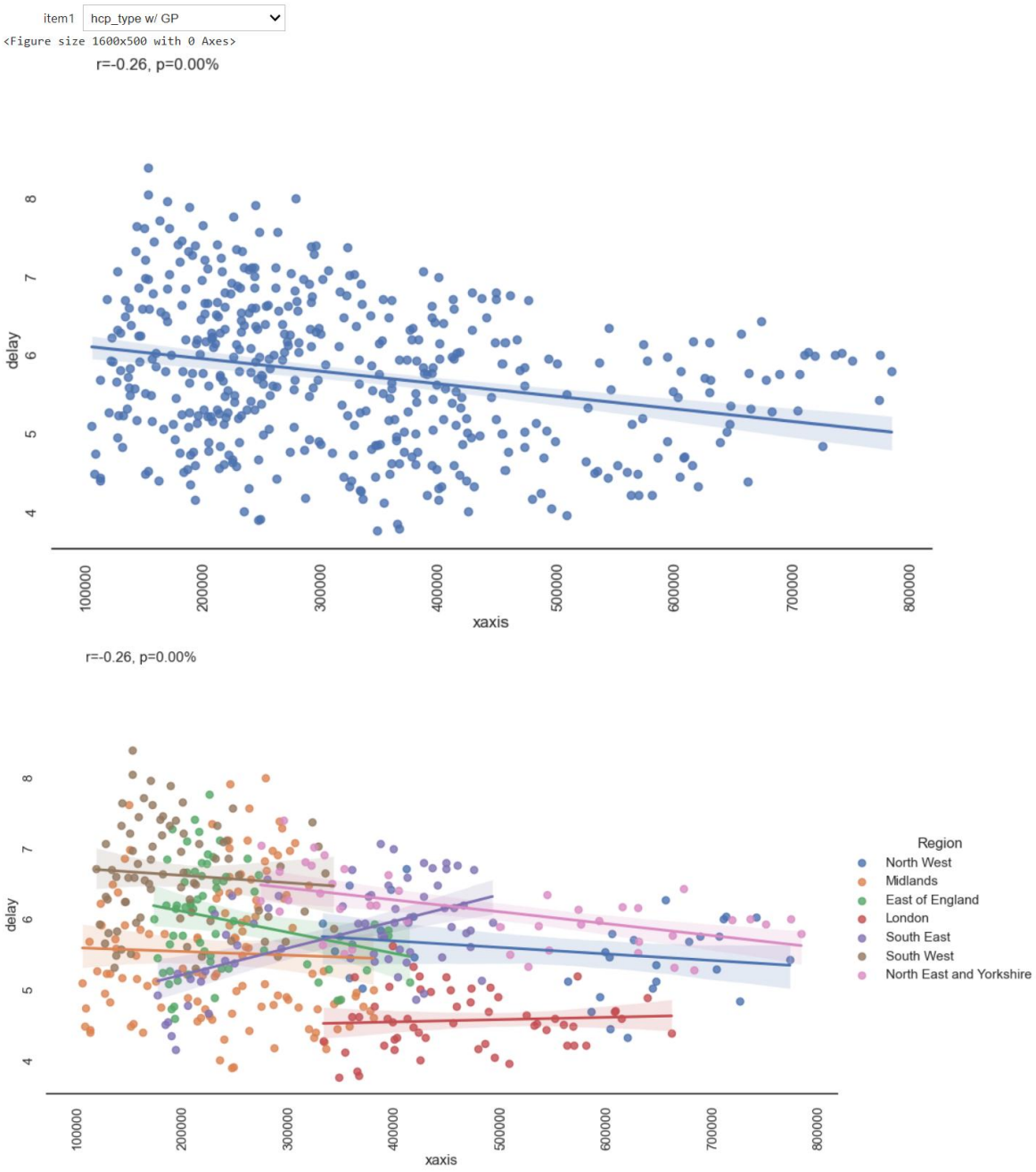
Mapping for Context Type	
Old Category	**New Category**
Care Related Encounter	Care Related Encounter
Unmapped	Bad mapping*
Inconsistent Mapping	Bad mapping*

Mapping for National Category	
Old Category	**New Category**
Planned Clinics	Planned Clinics*
Planned Clinical Procedure	Planned Clinics*
General Consultation Acute	General Consultation Acute
General Consultation Routine	General Consultation Routine
Clinical Triage	Clinical Triage
Unmapped	Bad mapping*
Inconsistent Mapping	Bad mapping*
Care Home Needs Assessment & ….	Other*
Patient contact during ….	Other*
Service provided by organisation …	Other*
Unplanned Clinical Activity	Other*
Social Prescribing Service	Other*
Non-contractual chargeable work	Other*
Group Consultation and Group Education	Other*
Structured Medication Review	Other*
Care Home Visit	Other*
Home Visit	Other*
Walk-in	Other*

Robustness Tool

Below is a picture of a tool that we used as a robustness too that looked at Delay to appointment and count of appointments of various types. This is done over ICBs and months. It can tell us if, for example, there is a relationship with number of appointments with a GP and a longer average delay to an appointment.

Figure 19 A Scatterplot Showing How Delay in Appointments and Count of Appointments of a Specific Type



We did not use it as the main body of our analysis, but it could be added in the future. Proper analysis would require multivariate analysis, with different types of appointments playing a part.

London Graphic

Figure 20 London Measures of Resources