

Case Study : Analysis of “OP Waiting List” datasets available at National Open Data Portal

-Poojitha Vishukumar

1.Introduction:

The report provides the knowledge of accessing and analysing the datasets available on open data platform. In this task, R program is developed to process OP waiting list data available on National Open Data portal (<http://data.gov.ie/>). This involves connecting to the CKAN platform and accessing the datasets and associated files for each year on the portal. The datasets from 2014 to 2018 is cleaned , integrated and analysed visually by aggregating the counts by hospital, speciality and age group .

2.Packages used:

- **ckanr**: ‘Client for the Comprehensive Knowledge Archive Network’ package is used to search, show and list the packages and resources in <https://data.gov.ie> API.
- **tidyverse**: is used to tidy the data, here separate function is used split and extract year from the date column.
- **dplyr**: is used for data manipulation such as filter, mutate, select and summarize is used for preprocessing of the data.
- **ggplot2**: is used to plot the various graph for visual analysis of data.

3. Challenges

3.1 Accessing the datasets available on Open Data Platform :

We have installed and loaded the ckanr package in order to setup and extract the resources from the desired OP waiting list package . ckanr is a full client for the CKAN API, wrapping all APIs, inclusive of reading and writing data. The url-<https://data.gov.ie> is setup by using ckanr setup function and resources (2014-2018 datasets) from OP waiting list packages is searched and retrieved as follows.

```
# Accessing the datasets available on Open Data Platform (www.data.gov.ie)

#setting up the url
ckanr_setup(url = "https://data.gov.ie")
# viewing all the available packages as table
package_list(as = "table")

#package retrieval
#Searching for the "OP waiting List" package
all_packages <- package_search(q = 'OP waiting List By Group Hospital')$results[1]

#Displaying all the resources available in OP waiting list package
all_resources <- package_show("7a6cc878-5b09-476e-be5e-6165a9d7130c", as="table")$resources

#Resources in Hospital package year wise
resource_2014 <- fetch(resource_show(id = "665c02c5-f01d-483f-a1ce-95290b186aaa", as = "table")$url)
resource_2015 <- fetch(resource_show(id = "2d91da47-6039-445e-b0d9-bde9ba1a600d", as = "table")$url)
resource_2016 <- fetch(resource_show(id = "a83d466f-2d12-4183-bdf1-e98afcc1d0d1", as = "table")$url)
resource_2017 <- fetch(resource_show(id = "4573fa36-4392-4b7b-9053-4a910d5c3257", as = "table")$url)
resource_2018 <- fetch(resource_show(id = "a7f03b43-008b-4174-9f3f-e08b37e11433", as = "table")$url)
```

3.2 Integrating the datasets into one dataset

The year-wise datasets are pre-processed by splitting the date column into corresponding year , selecting only the necessary useful informative columns such as Hospital name, Speciality, Age groups and few columns were renamed for uniformity and ease of merging. All the datasets were merged into a single dataframe and null values or black entries were processed with addition of new column,“Age_category” consisting of group name according to their age as child, adult and elderly people, respectively.

3.3 Aggregate counts

The OP waiting list data has been grouped into three main categories to calculate and study the aggregate count with respect to year, hospital , speciality and age group respectively. The resulting dataframe is thus, extracted as csv format file.

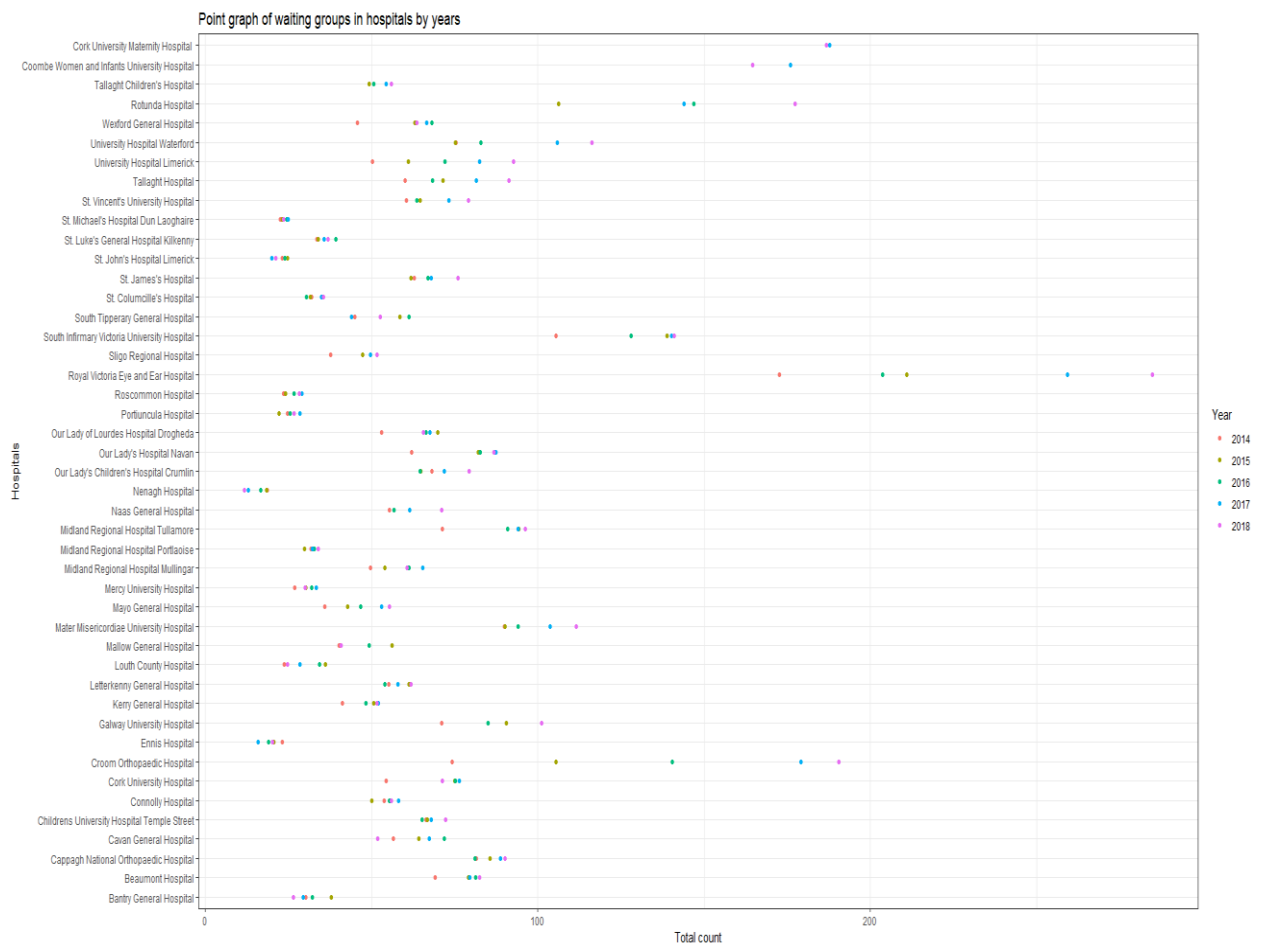
“Mean” function is used to get aggregated count of the following data for effective plotting of results.

The three summarized dataframes are:

- Hospital_df: provides the details of number of patients waiting on an average in each of the hospital across the years.
- Speciality_df: provides the details of number of patients waiting on an average for each specific speciality over the years.
- Age_df: provides the details of number of patients waiting on an average in each age group/ category over the years.

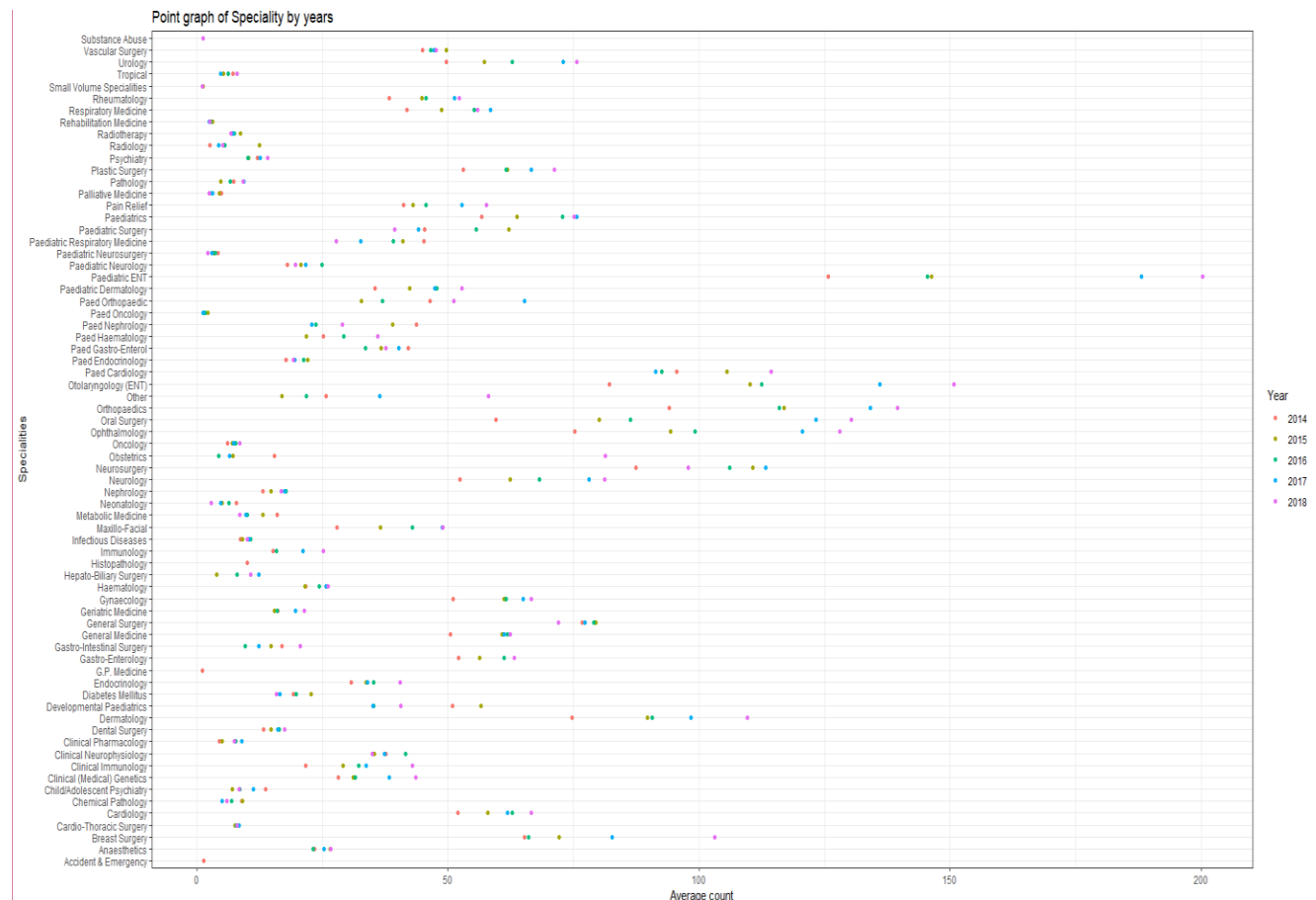
4. Visualization of the result:

4.1 Summarized data of waiting groups with respective to hospitals over the years



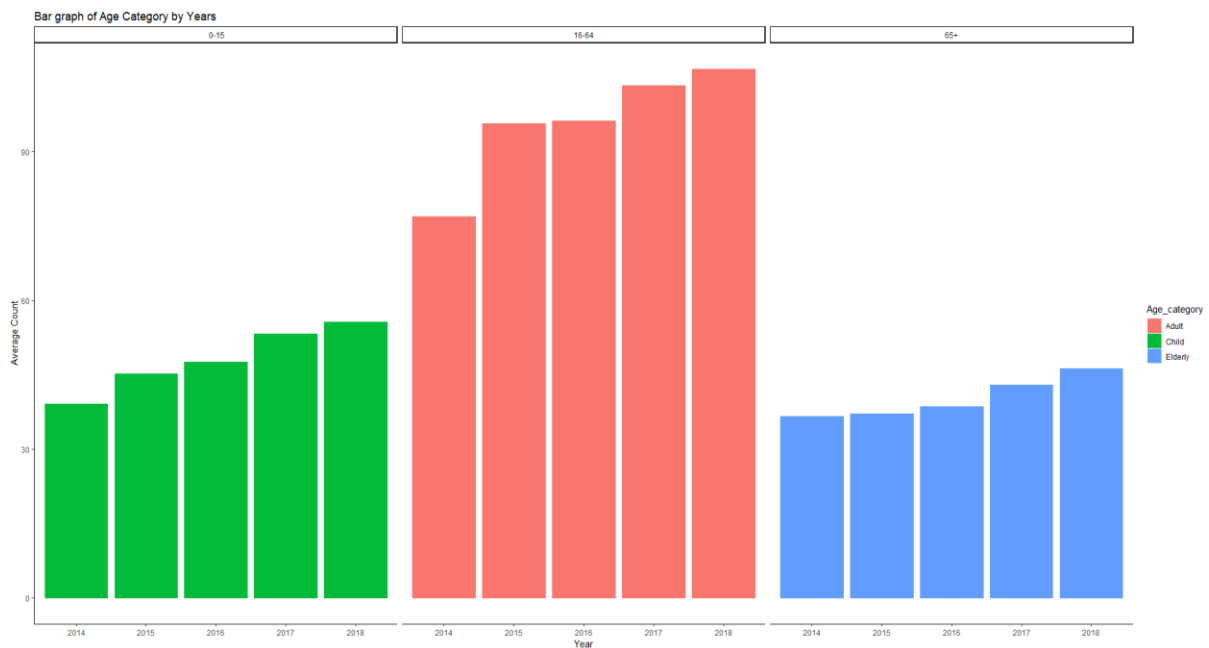
In the above point graph, we can observe that over the years, majority of hospitals have increased waiting count on an average except for the Cork University Maternity , Roscommon, Portlincula, Cork University, and Bantry General Hospital. Wherein the mean waiting count in 2018 has comparatively decreased when compared to previous years. Royal Victoria Eye and Ear hospital has the highest average waiting count with variance. On the other hand, Nenagh, Cavan and Bantry General Hospital have showed excellent improvement in reducing their average waiting time every year.

4.2 Summarized data of waiting groups with respective speciality across the years



The above point graph gives the visual of how many patients are waiting under each speciality over the years. Paediatric ENT has the highest average waiting count every year while rehabilitation medicine, neonatology, diabetes mellitus and paediatric respiratory medicine have showed improved waiting count by significantly decreasing the average waiting count ever year. On the other hand, specialities like Paed oncology, cardio-thoracic, infectious diseases, oncology are constant over the years and have minimum average waiting time comparatively.

4.3 Summarized data of waiting groups with respective to age group and category over the years



The above barchart depicts the summary of average waiting count of the different aged group patients over the years. A new column named age category was created classifying 0-15 years as child , 16-64 years as adult and 65+ years as Elderly people. Both age and category have been considered in the above graph for better interpretation. The graph has increasing slope for all the three age group for all the years depicting that the waiting time for all the age group has been increasing over the time. The average waiting time for elderly people is quite low than children and adults whereas it is the highest in adults. Also, in the year 2018, adults have waited the largest and in 2014, elderly have waited least.

5. Conclusion:

From this task, we can conclude that the waiting time has been increasing in majority of hospitals, with respect to speciality and different age groups over the span of years. Amongst hospitals, Royal Victoria Eye and Ear hospital, amongst specialities, Paediatric ENT and amongst age group, adults have the highest waiting time over the years. The ration of patients to respective staff and doctors for highest waiting time specialities can be maintained in order to reduce the average waiting time for the coming years.