

Is money omnipotent - Investigation of Canadian Health Care Budget's Impact on Waiting Time

Nammn Joshii

October 06, 2019

Contents

Introduction	1
Problem Statement	1
Data Resources	2
Data Model and Management	3
Downloading/Reading the Budget within the working directory:	3
Waiting Times for procedures in Canada by province. 2008 to 2018	4
Downloading/Reading the Waiting Time data within the working directory:	4
Constructing the Database	4
Data Cleansing	5
Postgresql Table Creation	8
Merging Budget1 and Waiting_Time Tables	8
Removing Unwanted Columns from Merged Table	8
Analysis	9
Analysis 1 : Budget vs Avg Waiting Time	9
Analysis 2 : Province vs Avg Waiting Time	9
Analysis 3 : Year vs Avg Waiting Time	10
Correlation between Waiting Time and Budget	11
Regression Analysis	12
Conclusion	12
Discussion: Negative or Positive Relationship?	13
References:	13

Introduction

Inez Rudderham, a 33 year-old woman from Nova Scotia, recently gained national attention. In a video on her Facebook page, she unfolds her heart-breaking story that she went undiagnosed with stage three anal cancer for two years, because of her lack of access to a family doctor. Unfortunately, this is not an isolated incident. In 2018, Canada's single-payer healthcare system has forced over a million patients to wait for essential medical treatments.

Problem Statement

The average treatment waiting time in Canada is 21 weeks after referral from general practitioner. A heart breaking revelation emerged after Fraser institute dug deeper into this matter, only to find that 72% of the morality cases due to such delays were actually preventable. Longer waiting time is a nationwide challenge for

the Canadian Healthcare system. It is a growing concern and has significant impacts on Canadian people's welfare. We truly believe this is a meaningful topic to investigate.

In this project, we will focus on the Canadian healthcare system. Specifically, the primary focus of our project is to investigate the relationship between the healthcare budget and waiting time.

It may seem trivial to assume the relationship is negative, but there are some other interesting questions we are trying to answer in this project. For example: How strong the correlation is? Is the relationship linear? Is there a marginal diminishing effect for waiting time when more and more budgets are put into the system?

The ultimate goal of our project is to provide remedial suggestions for both public and private healthcare participants in Canada to shorten the waiting time for medical services, and hopefully to promote welfare for the whole society .

Data Resources

We started our data search with following criteria: the availability of needed variables, the scope and time span of the databases, the quality of the data, and whether the source is trusted.

Specifically, in this project, we are investigating the relationship between the healthcare budgets and the waiting time for health care services. These two will be the independent and dependent variables in our study and are our main targets to search for.

In terms of the scope and time span, since we are interested in longer waiting time, a distinct issue of Canadian healthcare system, we limit our scope to the nation. Ideally, the dataset will have provincial data, so that we will be able to generate more insight by comparing data across Canada. For time span, our proposed time span of the study is five years. However, in the data searching process we tried to find the databases with the longest time span possible to give us more flexibility to trim and accommodate different time spans in the two variables. We will generate our final clean database using the overlapped portion of the data we collect.

Finally, about the data quality and trustworthiness of the data. We put our focus not only on the data itself, but also the support documents come with it. We carefully read the description files to find the definitions and methodologies conducted in the data collection process. We tried to exclude datasets with too much N/A values. We also take a look at the trustworthiness of the organization that provides the data.

Based on the above selection criteria, our final chosen data are collected from the Canadian Institute for Health Information (CIHI). Founded in 1994, CIHI is a not-for-profit organization that provides a good collection of Canadian healthcare system databases, measurements, and analyses report. The CIHI is also a reliable data source and is commonly used for research. In Josephs, K., & Fahey, J. (2009), the authors conclude that its databases are relatively accurate and compared favourably to other available data sources.

Installing Relevant Packages

We start by installing relevant packages which will help us in cleaning and analysing the data. They can be installed after removing the # symbol.

```
#install.packages("tidyverse")
#install.packages("RCurl")
#install.packages("readxl") # CRAN version
#install.packages("devtools")
#devtools::install_github("davidsjoberg/hablar")
```

Loading necessary libraries

After installing the packages, we load the relevant packages to better resolve the problem statement.

```
knitr::opts_chunk$set(echo = TRUE,
                      results = 'hide',
                      message = FALSE,
```

```
warnings = FALSE)
library("Hmisc")
library("tidyverse")
library(RPostgreSQL)
library(dplyr)
library(lubridate)
library(readr)
library(tidyr)
library(RCurl)
library(readxl)
library(hablar)
library(data.table)
library(zoo)
library(ggplot2)
library(tibble)
library(readxl)
```

National Health Expenditure Data, 1975 to 2018

Our primary independent variable is the medical budget. Our data resource is from the Canadian Institute for Health Information (CIHI). (<https://www.cihi.ca/sites/default/files/document/nhex-series-b-2018-en.xlsx>)

The data of total health expenditure by provinces/territories is the representative indicator of this factor, which includes total health expenditure, in millions of current dollars, for columns we have 10 provinces and 2 territories in Canada, and for column range from 1975 to 2018. (10 provinces and 3 territories are respectively Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick, Quebec, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia, Yukon, and Northwest Territories). Key significance of this dataset is the detailed total health expenditure 10 provinces. We use these data as the representative indicator of medical budget and then create the database.

Column	Meaning
year	Year of the Dataset (1975-2018)
[NL – Nvt.]	Province or territory (13 cols)
Values within each cell	Expenditure allocated

Data Model and Management

Downloading/Reading the Budget within the working directory:

```
if (!dir.exists("data/input/")) {
  if (!dir.exists("data")) {
    dir.create("data")
  }
  dir.create("data/input")
}

if (!file.exists("data/input/Budget.xlsx")) {
  download.file("https://www.cihi.ca/sites/default/files/document/nhex-series-b-2018-en.xlsx",
  destfile='data/input/Budget.xlsx', method='curl')
}
```

Waiting Times for procedures in Canada by province. 2008 to 2018

Our primary dependent variable is the Waiting Time data. Our data resource is from the Canadian Institute for Health Information (CIHI). (http://waittimes.cihi.ca/spreadsheets/Wait%20Times%20Export%20Data%202019_EN.xlsx)

This extensive waiting time data is categorized by provinces in Canada, Volumes of procedures affected by it and types of procedure. (Types of Procedures : Cataract, Bypass Surgery, Radiation Therapy, CT Scan, MRI Scan, Bladder Cancer Surgery, Breast Cancer Surgery, Colorectal Cancer Surgery, Lung Cancer Surgery, Prostate Cancer Surgery)

Key research significance of this data is that it is the primary dependent variable for the problem statement and it helps quantify the core problem.

Column	Meaning
Data Year	Year of the Dataset (2008-2018)
Province [NL – Nvt.]	Province or territory (13 cols)
Volume	Volume of Procedures
% Meeting Benchmark	% meeting Benchmark Waiting Time
50th Percentile (in Days)	Median Waiting Time
90th Percentile (in Days)	90th Percentile of Waiting Time

Downloading/Reading the Waiting Time data within the working directory:

```
if (!dir.exists("data/input/")) {  
  if (!dir.exists("data")) {  
    dir.create("data")  
  }  
  dir.create("data/input")  
}  
download.file("http://waittimes.cihi.ca/spreadsheets/Wait%20Times%20Export%20Data%202019_EN.xlsx",  
destfile='data/input/Waiting Times.xlsx', method='curl')
```

Constructing the Database

We create a PostgreSQL database named “Group Assignment” for our data. Both datasets are at year wise and the provincial level.

```
# Build will fail if the file isn't there.  
assertthat::assert_that(file.exists('db_connect.txt'),  
  msg = "Your connection file is missing.")  
  
con_file <- readr::read_lines('db_connect.txt')  
  
con <- RPostgreSQL::dbConnect(  
  PostgreSQL(),  
  host = con_file[1],  
  port = con_file[2],  
  user = con_file[3],  
  password = con_file[4],  
  dbname = con_file[5])
```

Data Cleansing

Initial Data Cleaning - Budget

In order to make the markdown file more scalable, we are using “if !file.exist” to check whether the file “Budget.rds” already exists within the output directory (using “file.exists” function) than we would read it into the RStudio (using “read_xlsx” function) or else we would use the code below it to cleanse the data.

We are endeavouring to clean data collected above using dplyr package. We are first reading the downloaded xlsx file and then creating a new subset table by selecting a range (A5:K4) from Total sheet of downloaded Budget dataset. We will consider budget time from 2008 onwards, therefore, eliminating unnecessary year rows. We then create a new Canada column which is sum total of all the provinces using “rowSums” function.

We then rename all the provinces into more readable format by converting the collected data into matrix (using as.tibble) and then use “names” function to rename the provinces. We also rename 2017 and 2018 year rows as it has alphabet “f” written against it and then we convert the column Year into numeric. Than we used “gather” function to develop key-value pairs between Province and Budget columns.

Then, we save this cleaned file (Budget1) into the output directory in RDS format. We included an error safe function “!file.exists” to crete the output directory if not done already.

```
if (!file.exists('data/output/Budget.rds')) {
  Budget <- read_xlsx("data/input/Budget.xlsx", range = "Total!A5:K49")
  Budget <- Budget[-c(1:33), ]
  Budget$Canada <- rowSums( Budget[,2:11])
  Budget <- as.tibble(Budget)
  names(Budget)[names(Budget) == "N.L."] <- "NL"
  names(Budget)[names(Budget) == "P.E.I."] <- "PEI"
  names(Budget)[names(Budget) == "N.S."] <- "NS"
  names(Budget)[names(Budget) == "N.B."] <- "NB"
  names(Budget)[names(Budget) == "Que."] <- "QC"
  names(Budget)[names(Budget) == "Ont."] <- "ON"
  names(Budget)[names(Budget) == "Man"] <- "MB"
  names(Budget)[names(Budget) == "Sask."] <- "SK"
  names(Budget)[names(Budget) == "Alta."] <- "AB"
  names(Budget)[names(Budget) == "B.C."] <- "BC"
  Budget$Year[Budget$Year %in% "2017 f"] <- "2017"
  Budget$Year[Budget$Year %in% "2018 f"] <- "2018"
  Budget$Year <- as.numeric(Budget$Year)
  Budget1 <- gather(Budget, key = "Province", value = "Budget", -Year)
  if (!file.exists('data/output')) {
    dir.create('data/output')
  }
  saveRDS(Budget1, 'data/output/Budget1.rds')
} else {
  Budget1 <- readRDS('data/output/Budget1.rds')
}
```

Initial Data Cleaning - Waiting Time

Waiting time dataset was the most interesting dataset we came across till date. It required non traditional approach for cleaning.

In order to make the markdown file more scalable, we are using “if !file.exist” to check whether the file “Waiting Time.rds” already exists within the output directory (using “file.exists” function) than we would read it into the RStudio (using “read_xlsx” function) or else we would use the code below it to cleanse the data.

We are endeavouring to clean data collected above using dplyr, Hmisc and tidyverse package. We are first reading the downloaded xlsx file and then creating a new subset table by selecting a range (A2:AY123) from Provincial Wait Times sheet of downloaded Waiting Time dataset. The entire data was in character format so had to convert all the values within columns into numeric using “as.numeric(as.character(x))” function.

The rows and columns with the original dataset was merged for multiple rows and columns. Thus, importing it caused most of the columns to change into NA values. Therefore, we used a while loop and then a for loop to repeat year values from previous row into the NA cells.

We will consider budget time from 2008 onwards, therefore, eliminating unnecessary year rows. We then create a new Canada column which is sum total of all the provinces using “rowSums” function.

We then created 3 new columns Total Volume, Total 50thP and Avg50thP for our further analysis using “rowSums” and “rowMeans” functions.

We then had to rename all the provinces into into more readable format by converting the collected data into matrix (using as.tibble) and then use “names” function to rename the provinces. The function “setNames” did work due to different value format, therefore that to change columns names individuals.

Then, we save this cleaned file (Waiting_Time) into the output directory in RDS format. We included an error safe function “!file.exists” to crete the output directory if not done already.

```
if (!file.exists('data/output/Waiting_Time.rds')) {

#Selecting Data

Waiting_Time <- read_xlsx("data/input/Waiting Times.xlsx",
                          range = "Provincial Wait Times!A2:AY123")

# Converting Char to Numeric

i = c(3:51)
Waiting_Time[, i] <- apply(Waiting_Time[, i], 2,
                          function(x) as.numeric(as.character(x)))

# Remove NA from year

i=0
Waiting_Time$Year <- Waiting_Time$`Data Year`
while (i<11) {
Waiting_Time$Year<-
  ifelse(is.na(Waiting_Time$`Data Year`), lag(Waiting_Time$Year1,1),
        Waiting_Time$`Data Year`)
i = i + 1
}

for(i in 1:nrow(Waiting_Time)){
  Waiting_Time$Year[i] <- ifelse(is.na(Waiting_Time$`Data Year`[i]), Waiting_Time$Year[i-1],
                                Waiting_Time$`Data Year`[i])
}

#Adding Columns

Waiting_Time$TotalVolume <-
  rowSums(Waiting_Time[,c(3,7,19,23,27,31,34,37,40,43,46,49)], na.rm = TRUE)

Waiting_Time$Total50thP <- rowSums(Waiting_Time[,c(5,9,17,21,25,29,32,35,38,41,44,47,50)], na.rm = TRUE)
```

```
Waiting_Time$Avg50thP <-
  rowMeans(Waiting_Time[,c(5,9,17,21,25,29,32,35,38,41,44,47,50)],
    na.rm = TRUE)
```

#Renaming Columns

```
Waiting_Time <- as.tibble(Waiting_Time)
names(Waiting_Time)[names(Waiting_Time) == "Volume...3"] <- "HR Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...4"] <- "HR%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...5"] <- "50thHR"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...6"] <- "90thHR"
names(Waiting_Time)[names(Waiting_Time) == "Volume...7"] <- "KR Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...8"] <- "KR%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...9"] <- "50thKR"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...10"] <- "90thKR"
names(Waiting_Time)[names(Waiting_Time) == "Volume...11"] <- "HFRS Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...12"] <- "HFRS%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Hours)...13"] <- "50thHFRS"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Hours)...14"] <- "90thHFRS"
names(Waiting_Time)[names(Waiting_Time) == "Volume...15"] <- "HFRE Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...16"] <- "HFRE%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Hours)...17"] <- "50thHFRE"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Hours)...18"] <- "90thHFRE"
names(Waiting_Time)[names(Waiting_Time) == "Volume...19"] <- "Cat Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...20"] <- "Cat%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...21"] <- "50thCat"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...22"] <- "90thCat"
names(Waiting_Time)[names(Waiting_Time) == "Volume...23"] <- "BS Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...24"] <- "BS%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...25"] <- "50thBS"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...26"] <- "90thBS"
names(Waiting_Time)[names(Waiting_Time) == "Volume...27"] <- "RT Volume"
names(Waiting_Time)[names(Waiting_Time) == "% Meeting Benchmark...28"] <- "RT%MB"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...29"] <- "50thRT"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...30"] <- "90thRT"
names(Waiting_Time)[names(Waiting_Time) == "Volume...31"] <- "CT Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...32"] <- "50thCT"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...33"] <- "90thCT"
names(Waiting_Time)[names(Waiting_Time) == "Volume...34"] <- "MRI Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...35"] <- "50thMRI"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...36"] <- "90thMRI"
names(Waiting_Time)[names(Waiting_Time) == "Volume...37"] <- "BCS Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...38"] <- "50thBCS"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...39"] <- "90thBCS"
names(Waiting_Time)[names(Waiting_Time) == "Volume...40"] <- "BRC Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...41"] <- "50thBRC"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...42"] <- "90thBRC"
names(Waiting_Time)[names(Waiting_Time) == "Volume...43"] <- "CCS Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...44"] <- "50thCCS"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...45"] <- "90thCCS"
names(Waiting_Time)[names(Waiting_Time) == "Volume...46"] <- "LCS Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...47"] <- "50thLCS"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...48"] <- "90thLCS"
```

```

names(Waiting_Time)[names(Waiting_Time) == "Volume...49"] <- "PCS Volume"
names(Waiting_Time)[names(Waiting_Time) == "50th Percentile (in Days)...50"] <- "50thPCS"
names(Waiting_Time)[names(Waiting_Time) == "90th Percentile (in Days)...51"] <- "90thPCS"

if (!file.exists('data/output')) {
  dir.create('data/output')
}
saveRDS(Waiting_Time, 'data/output/Waiting_Time.rds')
} else {
  Waiting_Time <- readRDS('data/output/Waiting_Time.rds')
}

```

Postgresql Table Creation

When then pushed data into the database. The `post_data` function failed to CASCADE each time we ran it. So had to use the tradition `dbWriteTable` function. This ensures we could run the script multiple times without errors.

```

dbWriteTable(con,
             "Budget1",
             Budget1,
             row.names = FALSE,
             overwrite = TRUE)

dbWriteTable(con,
             "Waiting Time",
             Waiting_Time,
             row.names = FALSE,
             overwrite = TRUE)

```

Merging Budget1 and Waiting_Time Tables

Now that we have cleansed and aligned columns of both the datasets into a database table, we merge both the tables into a new table called Merged using “`dbGetQuery`” and “INNER JOIN” SQL functions.

```

Merged <- dbGetQuery(con,
  'SELECT *
  FROM "Budget1" AS B
  INNER JOIN "Waiting Time" w ON B."Year" = W."Year" AND B."Province" = W."Province"')

```

Removing Unwanted Columns from Merged Table

We subset Merged table into a new table called as Merged_Final such that the final data we have has relevant data from 2013 to 2018 for Budget, Total Volume, Total 50thP and Avg50thP columns. We filterout all the Canada (!=) rows from Province Column as it will skew the distribution (using “`filter`” function). The data table Merged_Final is now ready for analysis.

```

Merged <- subset(Merged, select = -c(4:55))
Merged <- subset(Merged, Merged$Year >= 2013 & Merged$Year <= 2018)
Merged_Final = Merged %>%
  filter(Province != "Canada")

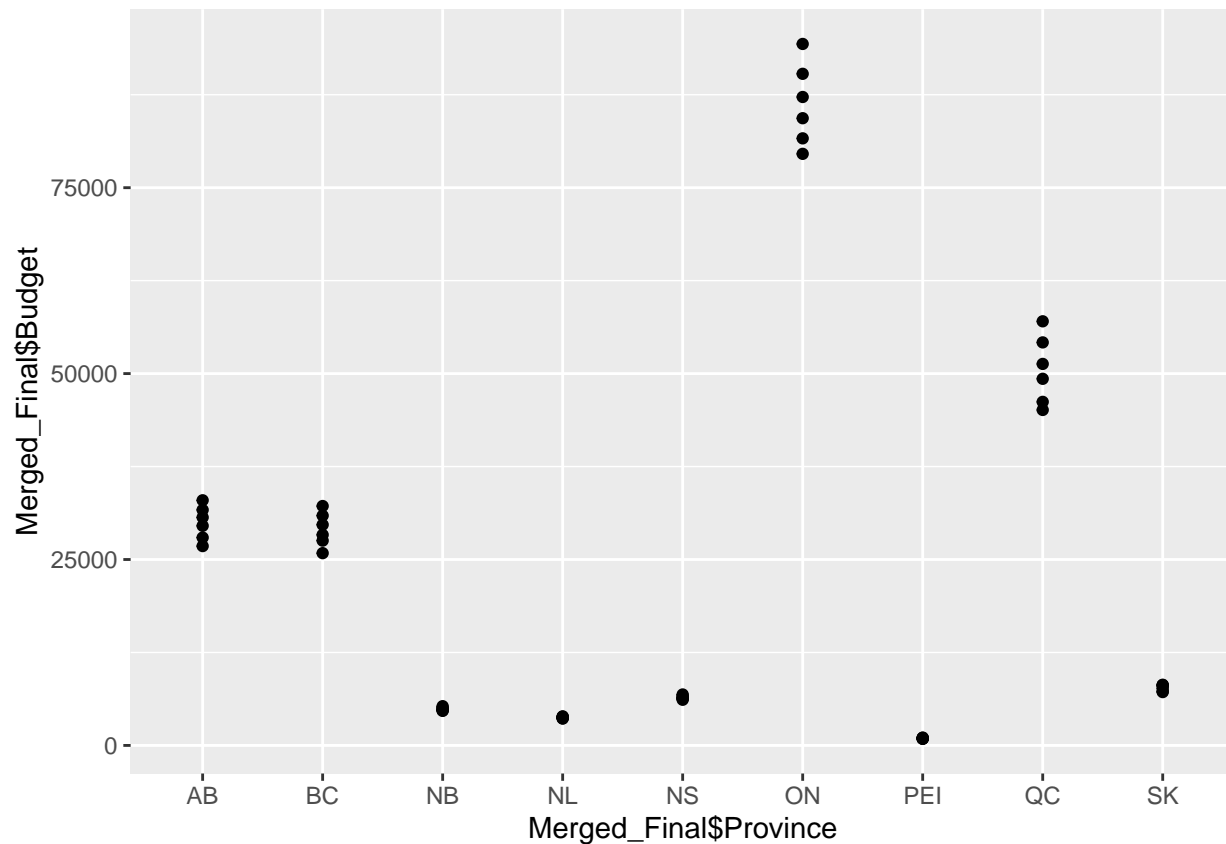
```


Analysis

Analysis 1 : Budget vs Avg Waiting Time

In order to provide actionable insights, we plot interaction between Province & Budget using qplot package.

```
qplot(Merged_Final$Province, Merged_Final$Budget, data = Merged_Final)
```

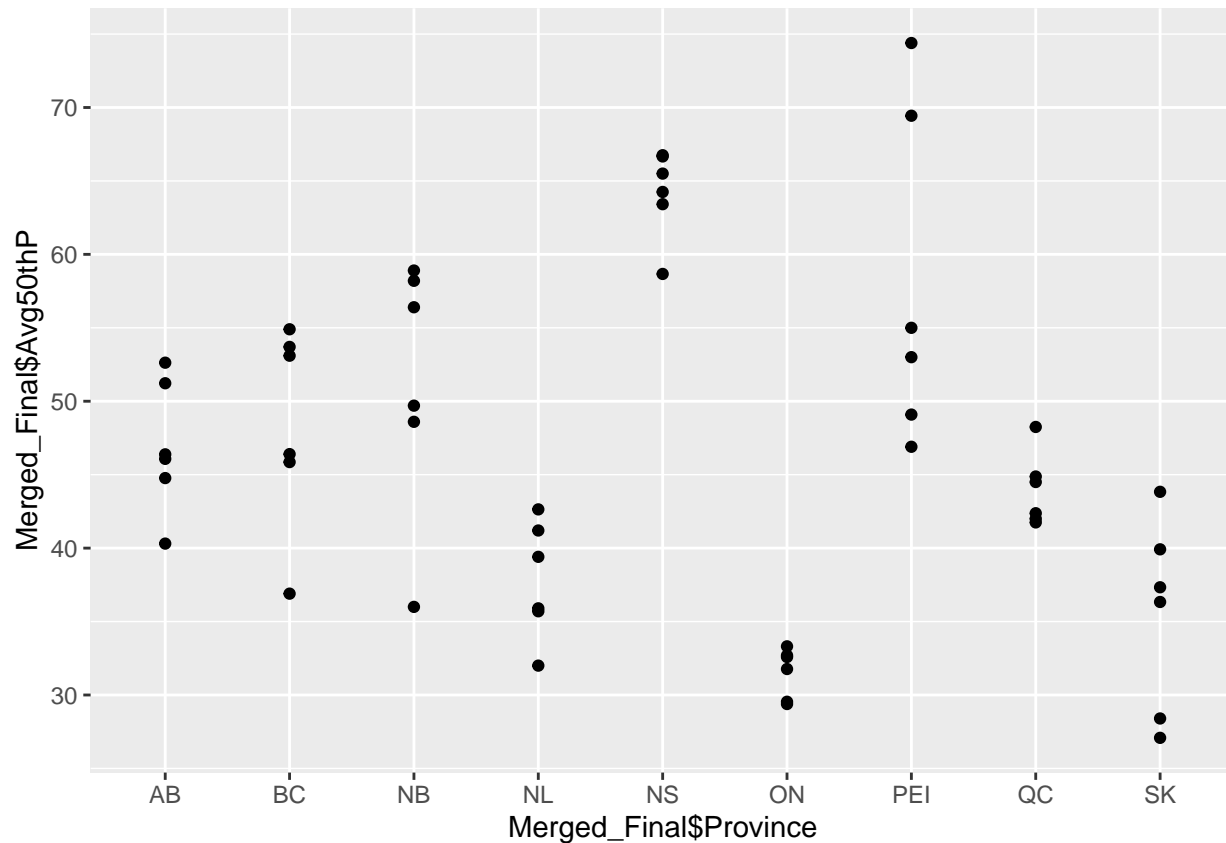


From the above interaction, we can establish that the amount of medical finances differs greatly depending on which province we analyze. An interesting fact is that the increase in budget amount is rather stable for each region.

Analysis 2 : Province vs Avg Waiting Time

Now let's see the interaction between Province & Avg Waiting Time when we plot it using qplot.

```
qplot(Merged_Final$Province, Merged_Final$Avg50thP, data = Merged_Final)
```

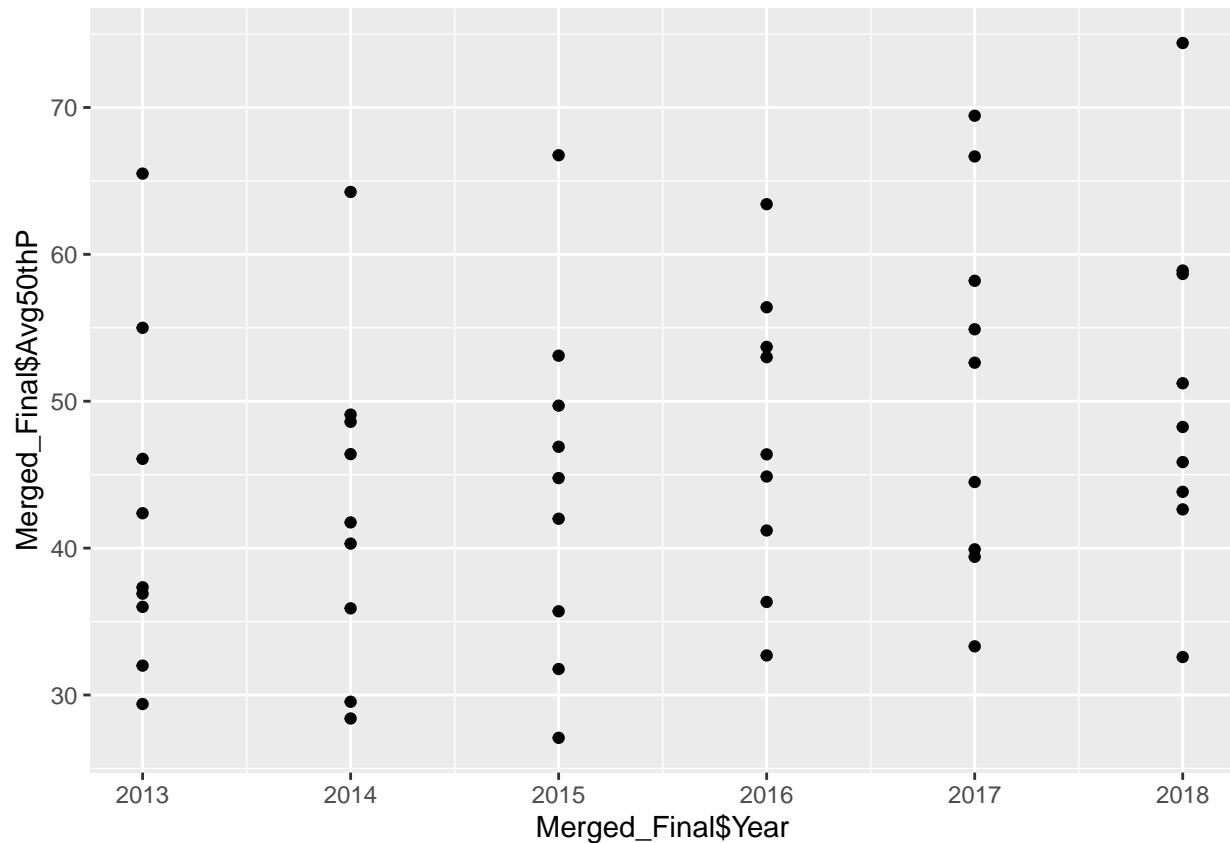


The graph above establishes that the increase in waiting time is not a regular trend in Canada, but rather a feature in several provinces. Different provinces have different waiting time trends and the largest deviation within years is spotted in British Columbia, New Brunswick and Prince Edward Island.

Analysis 3 : Year vs Avg Waiting Time

Finally, let us understand the interaction Year & Avg Waiting Times using `qplot`.

```
qplot(Merged_Final$Year, Merged_Final$Avg50thP, data = Merged_Final)
```

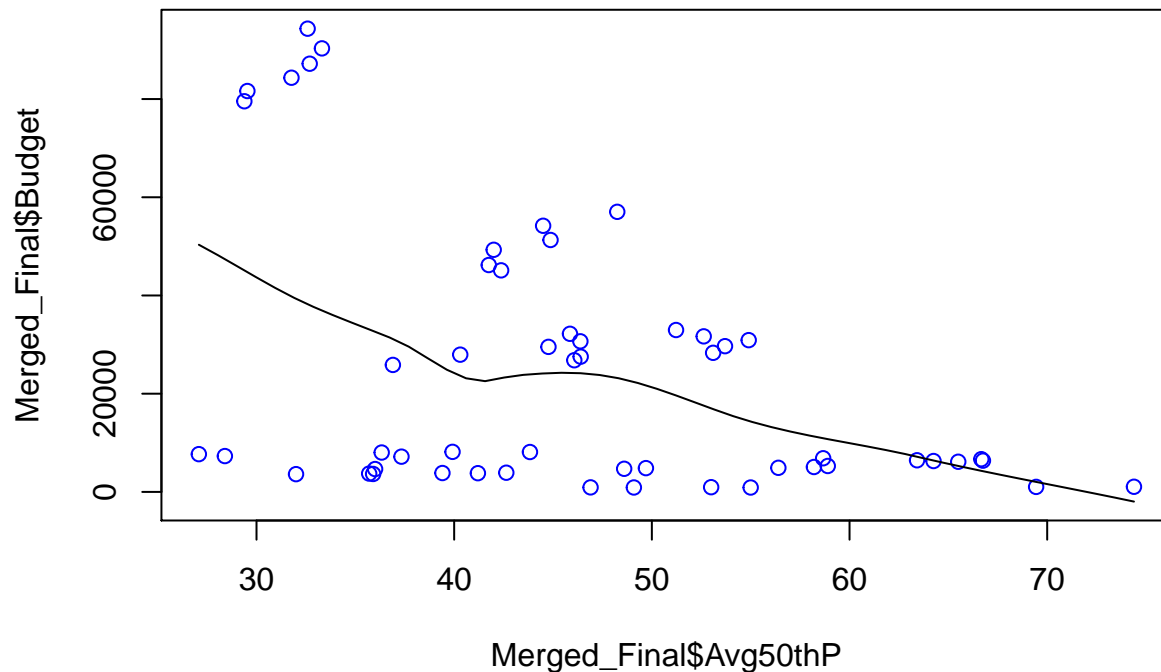


From this graph, we can suggest that, overall, each province experienced an increase in waiting time in 2018 compared to 2013. However, there is no stable percent of increase annually, meaning that a specific trend could affect the amount of waiting time in a particular year of observation and in a particular region.

Correlation between Waiting Time and Budget

We test the correlation between Waiting Time and Budget using “cor.test” function and then plot a scatter diagram using “scatter.smooth” function.

```
cor.test(Merged_Final$Avg50thP, Merged_Final$Budget)
scatter.smooth(Merged_Final$Avg50thP, Merged_Final$Budget, col='blue')
```



Pearson's product-moment correlation:

data: Merged_FinalAvg50thP and Merged_FinalBudget $t = -4$, $df = 52$, $p\text{-value} = 0.0006$ alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: -0.6 -0.2 sample estimates: cor -0.5

The graph above provides an interesting finding that there is a significant negative correlation between waiting time and healthcare medical budget (appr coef -0.5). Whenever the budget increases for a million Canadian dollar, the amount of waiting time, on average, decreases by 0.5 days.

The distribution within the graph can give us a clue that a particular trend (e.g. the dots grouped on the upper left corner) exists within a particular group of observations.

Regression Analysis

To understand Regression between Budget and Waiting Time we used "lm" function and then "summary" function to review its results.

```
lmWaitingTime = lm(Avg50thP~Budget, data = Merged_Final) #Create the linear regression
summary(lmWaitingTime) #Review the results
```

Call: `lm(formula = Avg50thP ~ Budget, data = Merged_Final)`

Residuals: Min 1Q Median 3Q Max -22.541 -6.259 -0.196 7.578 23.487

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 51.1050302 1.9141095 26.70 < 2e-16 **Budget -0.0001926 0.0000526 -3.66 0.00059** — Signif. codes: 0 '0.001' '0.01' '0.05' '0.1' '1'

Residual standard error: 10 on 52 degrees of freedom Multiple R-squared: 0.205, Adjusted R-squared: 0.19 F-statistic: 13.4 on 1 and 52 DF, p-value: 0.000585

Conclusion

In conclusion, we can establish that, although medical budget has an impact on the waiting time for patients in Canada, it seems that another major trigger can influence the amount of days that patients have to wait

for their treatments. In terms of medical budget, we found that there is a significant negative correlation between waiting time and healthcare medical finances (appr -0.5).

By grouping the data by provinces, we established that there is an effect of the amount of healthcare budget on the waiting time, though it should be noted that other possible trends and factors may bear influence on the waiting time amount. Based on the provided graphs, we can assume that a possible phenomenon can exist only within certain province, which can lead to a severe increase of the waiting time in healthcare system. A possible trend could be the aging population, lack of medical equipment of specialists required in the given region.

Discussion: Negative or Positive Relationship?

In the first iteration of our analysis, we did a regression analysis for all the data points in our database. We found a significant negative correlation between budget and waiting time. The result is consistent with our initial assumption.

However, when we plot our data points by province, the results surprised us: the provincial data shows a weak positive relation between the budget and the waiting time- a very counterintuitive result. Does it mean that the more budget is causing longer waiting time in Canada?

We reexamined our original model again, and found some possible issues about this result: 1. Our original model (regression) is oversimplified. It considers only one independent variable- the budget, without any control variables. This may lead to a problem that there are too many external factors (especially for the ones come from the inherent differences among provinces, for example, the total population, density, aging of the population, etc). Therefore, it does not make sense to include all the data points across Canada in our regression model. Plotting a line for each province and one for Canada nationwide respectively could give us more accurate information about the relationship between budget and waiting time.

2. Correlation does not necessarily mean causation. Since we did not include any econometric treatment to eliminate the simultaneity between the independent variable and dependent variable (e.g. a time lag). The simple regression analysis we ran can only show correlation, but does not have the ability to show the causality.

After considering the two issues above, we propose there might be a reverse causality exists, that means, it is not that budget affect waiting time. It is the longer waiting time forcing the Canadian government to put more and more budget in its healthcare system. However, unfortunately, due to some other reasons (facilities availability, more patents due to the aging population), the waiting time is still rising across Canada. This theory, however, requires further research of the possible outcomes and correlations.

The above discussion also gave us some directions for the potential future extensions of the Study. For example, more control variables can be in the model to decrease the impact from the externalities. Also, a time lag method could be beneficial to reveal the causality between the independent and dependent variable.

Thank You

References:

1. National Health Expenditure Trends, 1975 to 2018, CIHI (Nov 2018)
2. Wait times for priority procedures in Canada, CIHI (Mar 2019)