STATS 765 Project Milestone 2

Measuring the Impact of the 2021 Residential Tenancies Amendment Act

Chase Robertson - crob873

Goal

The goal of this project is to investigate the impact of the February 2021 Residential Tenancies Amendment Bill by observing changing rental trends on the New Zealand housing market.

Data Source

The data is sourced from the NZ Tenancy Services website. Three CSV files are provided with territorial, regional, and quarterly aggregate bond statistics. In order to translate Location Id's from the quarterly dataset into location names, it is also necessary to retrieve data from the Stats NZ site.

Data Processing

One of the first data processing tasks is to ensure each data column is imported as the correct type, to ensure proper usage later in the analysis.

```
# import quarterly dataset, show names and tabulate types
quarterly <- read.csv("Detailed Quarterly Tenancy.csv")
names(quarterly)
```

```
##  [1] "Time.Frame"               "Location.Id"
##  [3] "Dwelling.Type"            "Number.Of.Beds"
##  [5] "Total.Bonds"              "Active.Bonds"
##  [7] "Closed.Bonds"             "Median.Rent"
##  [9] "Geometric.Mean.Rent"      "Upper.Quartile.Rent"
## [11] "Lower.Quartile.Rent"      "Log.Std.Dev.Weekly.Rent"
```

```
table(sapply(quarterly, class))
```

```
##
## character   integer   numeric
##        10         1         1
```

Some of the ".Bond" and ".Rent" columns of each dataset import as character-type columns, which does not match their numeric contents. I found that removing the thousands-delimiting commas made converting to integer type possible.

The next data processing step is to convert the `Time Frame` column into more useful `year` and `quarter` columns. The `lubridate` library makes that a simple task.

I also joined the Stats NZ location ID and location name columns with the quarterly report, so that location names could be easily included in future analysis. It was also necessary to manually add ID `-99` to get the location name `"ALL"` to join correctly. That code and any other not visible here can be seen in the Appendix.

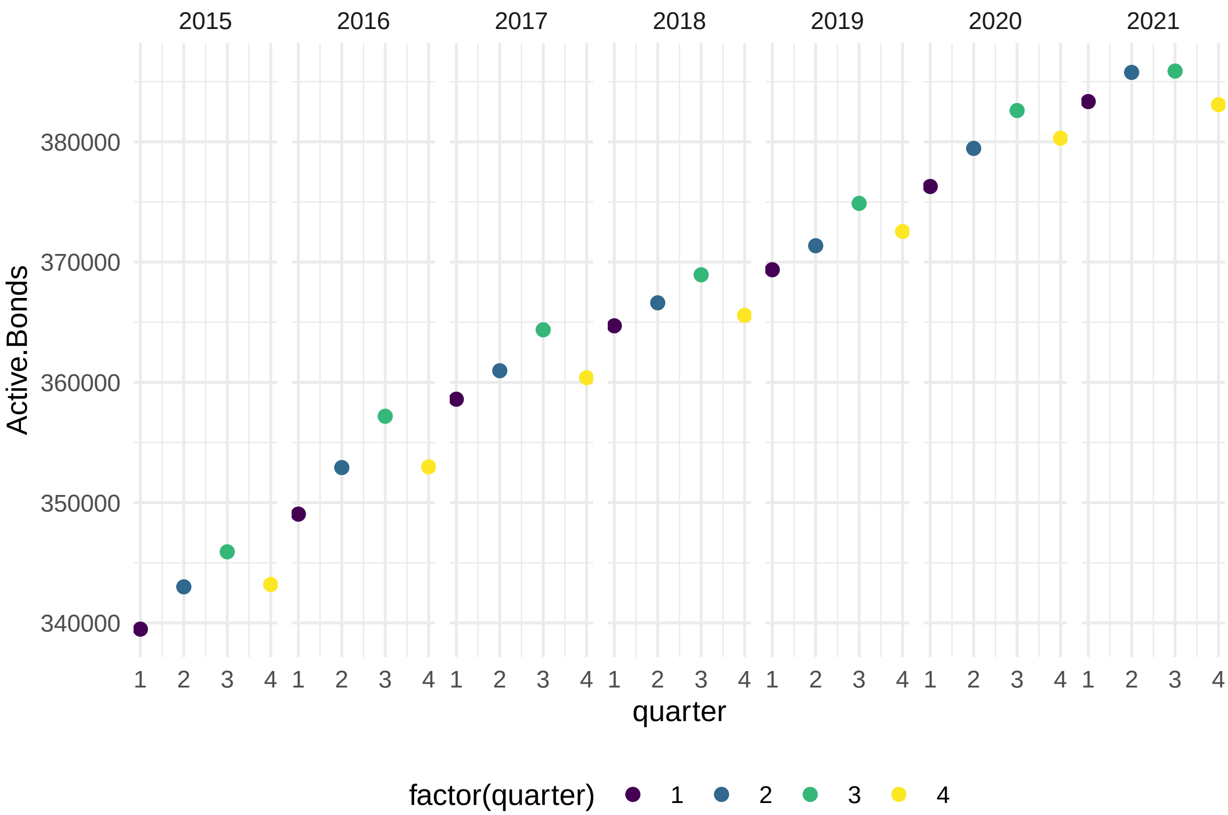
```
length(names(q.df))
```

```
## [1] 16
```

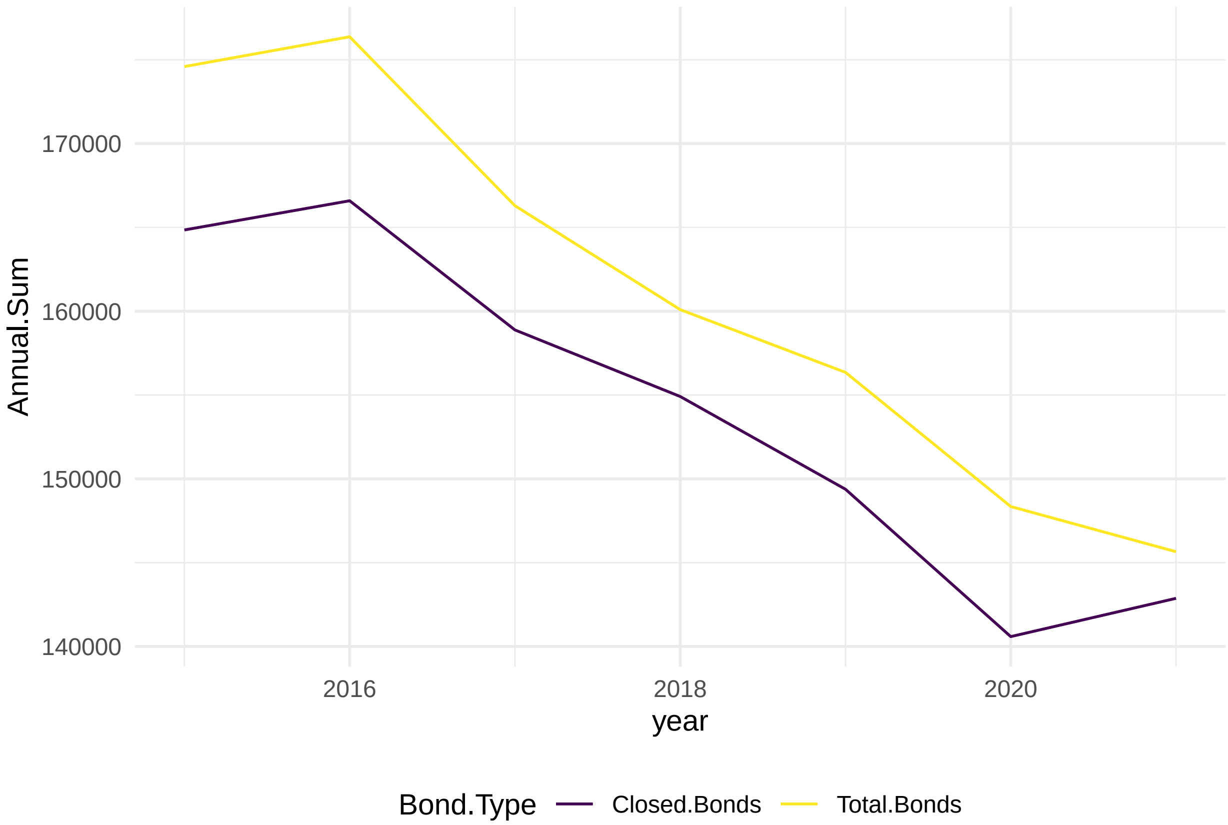
A similar set of steps was followed to import the regional and territorial datasets as well, as can also be seen in the Appendix.

Data Exploration

The first area of the data explored is the aggregate active bond numbers before and after Bill implementation. The question to keep in mind is: are there any noticeable patterns distinguishing the time before Feb 2021 from the time after?

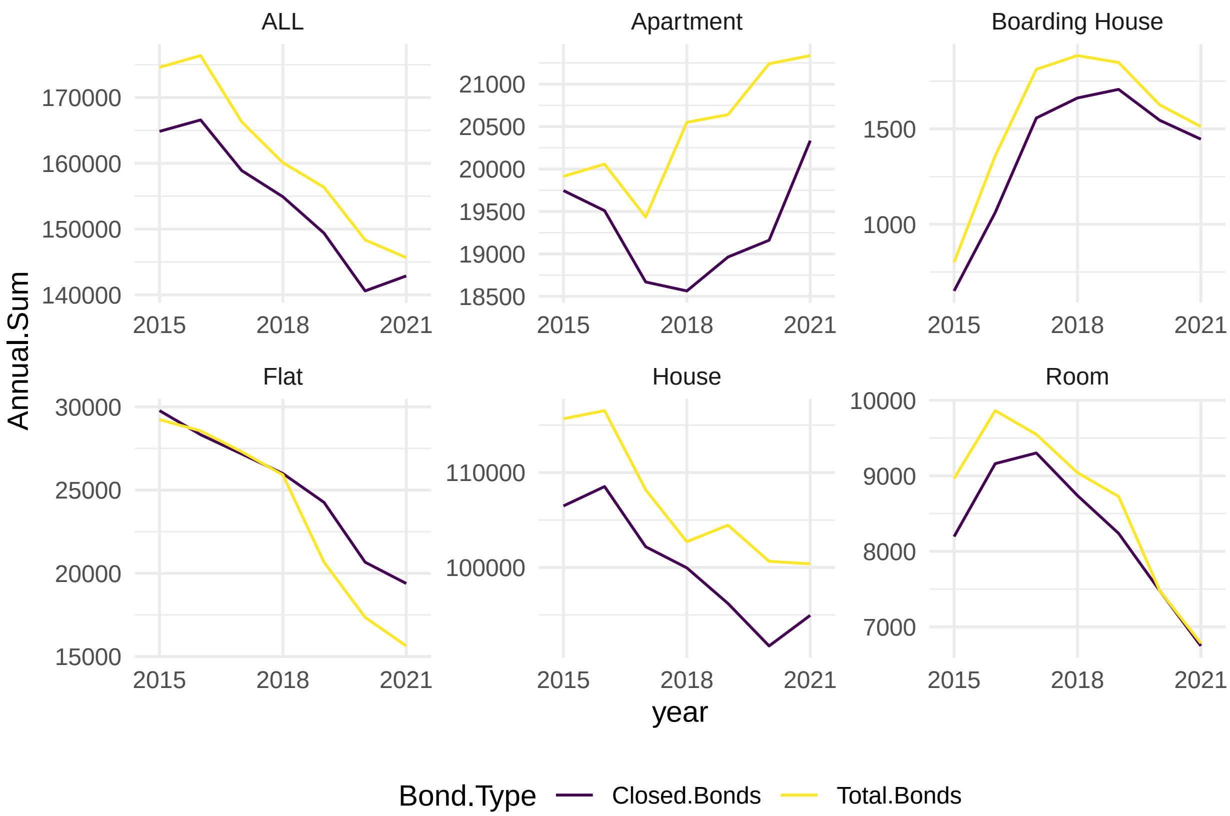


There doesn't seem to be any noticable change in the number of Active Bonds before vs after February 2021 on a quarterly basis. This suggest no universal impact from the Bill. But what about the number of Total or Closed Bonds?

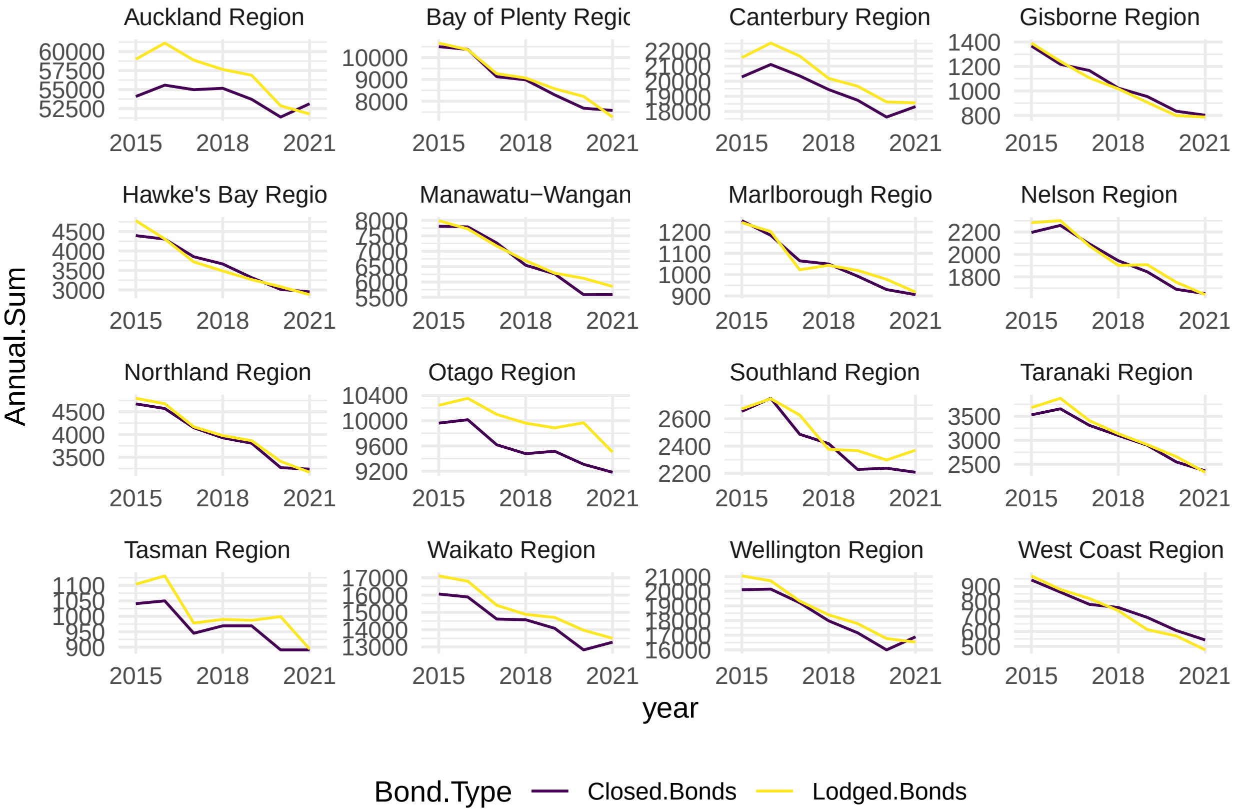


It seems that there may have been some unexpected increase in annual sum of Closed Bonds between 2020 and 2021. The opposite pattern can be seen between 2015 and 2016, however, so it may be a natural market correction.

I also wonder: is that 2021 increase reflected across all types of dwellings?



It seems that houses and apartments are the main contributors to the increase, and they are the more common dwelling types in the database. Now we can analyse the monthly regional data to check if any locations were more or less affected.



Auckland, Waikato, Wellington, and Canterbury demonstrate the upward trend in closed bonds. Those regions also harbor the largest cities in New Zealand, which begs the question: is this a pattern exclusive to urban environments? Could this be an "urban flight" phenomenon caused by COVID-19, rather than the Bill?

Analytical Plan

Based on the information illuminated in the data exploration, it seems that the effects of the Bill, if any, have varied across different dwelling types and locations. One potential path of analysis is constructing a paired t-test to see if there is a statistically significant difference between those bond measures in the time periods before and after the Bill. Perhaps a better way to test that hypothesis is to attempt to infer causality with a Bayesian structural time-series model. By building a Bayesian model from the data before the Bill's implementation, we can attempt to predict the data after the Bill, and compare those predictions with the actual data to infer the Bill's effect.

Appendix

Appendix includes all code necessary to replicate, plus some extra graphs not included in the written analysis.

```
# import quarterly dataset, show names and tabulate types
quarterly <- read.csv("Detailed Quarterly Tenancy.csv")
names(quarterly)
table(sapply(quarterly, class))

# remove comma from numeric values and convert columns
library(tidyverse)
q2 <- quarterly %>%
```

```r
    mutate(across(str_subset(names(quarterly),
                             regex(".Bond|(n|e).Rent|.Id")),
                  function(x) {as.integer(gsub(',', '', x))}))

# create year and month columns from Time.Frame
library(lubridate)
q3 <- q2 %>%
    mutate(date = mdy_hm(Time.Frame)) %>%
    mutate(year = year(date)) %>%
    mutate(quarter = quarter(date)) %>%
    mutate(date = paste(year, quarter, sep='-'))

# import and join location names
areas <- read.csv("statsnzstatistical-area-2-2019-centroid-true-CSV/statistical-area-2-2019-centroid-tru
locations <- areas %>%
    mutate(ID = SA22019_V1_00) %>%
    mutate(Location = SA22019_V1_00_NAME) %>%
    select(c(ID, Location))
locations <- rbind(locations, list(-99, 'ALL'))
q.df <- left_join(q3, locations, by=c('Location.Id' = 'ID'))
length(names(q.df))

# import and clean regional dataset
regional <- read.csv("rentalbond-data-regional.csv")
r2 <- regional %>%
    mutate(across(c(Lodged.Bonds, Closed.Bonds),
                  function(x) {as.integer(gsub(',', '', x))}))
r3 <- r2 %>%
    mutate(date = ymd(Time.Frame)) %>%
    mutate(year = year(date)) %>%
    mutate(month = month(date)) %>%
    select(-Time.Frame)
r.df <- r3

#import and clean territorial dataset
tla <- read.csv("rentalbond-data-tla.csv")

t2 <- tla %>%
    mutate(across(c(Lodged.Bonds, Active.Bonds),
                  function(x) {as.integer(gsub(',', '', x))}))
t3 <- t2 %>%
    mutate(date = ymd(Time.Frame)) %>%
    mutate(year = year(date)) %>%
    mutate(month = month(date)) %>%
    select(-c(Time.Frame, date))
t.df <- t3

# prepare for plotting
library(ggplot2)
theme_set(theme_minimal())
library(viridis)

# not pictured - quarterly Total Bonds (very skewed toward lockdowns)
```

```r
q.df %>%
    filter(Location == 'ALL', Dwelling.Type == 'ALL',
           Number.Of.Beds == 'ALL', year > 2014) %>%
    ggplot(aes(x=quarter, y=Total.Bonds, colour=factor(quarter))) +
    facet_grid(~year) + geom_point(size=2) +
    scale_color_viridis(discrete=T) + theme_minimal()

# quarterly Active Bonds
q.df %>%
    filter(Location == 'ALL', Dwelling.Type == 'ALL',
           Number.Of.Beds == 'ALL', year > 2014) %>%
    ggplot(aes(x=quarter, y=Active.Bonds, colour=factor(quarter))) +
    facet_grid(~year) + geom_point(size=2) +
    scale_color_viridis(discrete=T) + theme_minimal() +
    theme(legend.position = "bottom")

# not pictured - Closed Bonds (skewed by lockdowns)
q.df %>%
    filter(Location == 'ALL', Dwelling.Type == 'ALL',
           Number.Of.Beds == 'ALL', year > 2014) %>%
    ggplot(aes(x=quarter, y=Closed.Bonds, colour=factor(quarter))) +
    facet_grid(~year) + geom_point(size=2) +
    scale_color_viridis(discrete=T) + theme_minimal()

# Total and closed annual bonds per year
q.type = q.df %>%
    filter(Location == 'ALL', Dwelling.Type == 'ALL',
           Number.Of.Beds == 'ALL', year > 2014) %>%
    pivot_longer(c(Total.Bonds, Closed.Bonds),
                 names_to="Bond.Type", values_to="Bond.Number") %>%
    group_by(year, Bond.Type) %>%
    summarise(Annual.Sum = sum(Bond.Number))

q.type %>%
    ggplot(aes(x=year, y=Annual.Sum, colour=Bond.Type)) +
    geom_line() +
    scale_colour_viridis(discrete=T) +
    theme_minimal() +
    theme(legend.position = "bottom")

# total and closed annual bonds per dwelling type
q.building = q.df %>%
    filter(Location == 'ALL', Number.Of.Beds == 'ALL', year > 2014) %>%
    pivot_longer(c(Total.Bonds, Closed.Bonds),
                 names_to="Bond.Type", values_to="Bond.Number") %>%
    group_by(year, Bond.Type, Dwelling.Type) %>%
    summarise(Annual.Sum = sum(Bond.Number))

q.building %>%
    ggplot(aes(x=year, y=Annual.Sum, colour=Bond.Type)) +
    facet_wrap(~Dwelling.Type, scales="free") +
    geom_line() +
    scale_colour_viridis(discrete=T) +
```

```r
    theme_minimal() +
    theme(legend.position = "bottom") +
    scale_x_discrete(name="year",limits=c(2015,2018,2021))

# lodged and closed bonds by region
r.loc = r.df %>%
    filter(Location.Id > 0, year > 2014, year < 2022) %>%
    pivot_longer(c(Lodged.Bonds, Closed.Bonds),
                 names_to="Bond.Type", values_to="Bond.Number") %>%
    group_by(year, Bond.Type, Location) %>%
    summarise(Annual.Sum = sum(Bond.Number))

r.loc %>%
    ggplot(aes(x=year, y=Annual.Sum, colour=Bond.Type)) +
    facet_wrap(~Location, scales="free") +
    geom_line() +
    scale_colour_viridis(discrete=T) +
    theme_minimal() +
    theme(strip.text.x = element_text(hjust = 0.1),
          legend.position = "bottom") +
    scale_x_discrete(name="year",limits=c(2015,2018,2021))

# rent quartiles by region
r.time = r.df %>%
    filter(Location.Id > 0, year > 2014, year < 2022) %>%
    pivot_longer(c(Median.Rent, Upper.Quartile.Rent, Lower.Quartile.Rent),
                 names_to="Type", values_to="Value") %>%
    group_by(year, Type, Location) %>%
    summarise(Annual.Mean = mean(Value))

r.time %>%
    ggplot(aes(x=year, y=Annual.Mean, colour=Type)) +
    facet_wrap(~Location, scales="free") +
    geom_line() +
    scale_colour_viridis(discrete=T) +
    theme_minimal() +
    theme(strip.text.x = element_text(hjust = 0.1),
          legend.position = "bottom") +
    scale_x_discrete(name="year",limits=c(2015,2018,2021))
```