Fingal, County Dublin Tree Project

Introduction

The aim of the project is to look at street tree coverage in Fingal, County Dublin and compare the incidence of tree coverage with the age of residents, family makeup, population size and house prices.

I chose this project because of an interest in trees and their impact on community wellbeing (Coillte, 2020), (The Royal Parks, 2020). I focused on neighbourhood demographics to strengthen the case for increased tree coverage based on the belief that trees help communities.

Data Sources

Trees

I sourced tree data from Fingal County Council publicly available open data (Smart Dublin - Dublinked, 2020). The tree dataset had approx. 36k records (5 Mb). The dataset included the location address, neighbourhood, and other various trees details of surveyed trees. The neighbourhood details were provided in a field labelled "Town" which more or less corresponded with the national electoral division data (CSO, 2020). Before loading the dataset into R, I completed some data cleansing in Excel. I reduced some town names which contained North, South, East and West to a single town name. This was the case with Swords.

Data Cleaning

The project was carried out in R. I amalgamated town names into neighbouring towns when the town name didn't correspond with an electoral division. It was necessary for town names to correspond with electoral divisions as the demographic information I was using came from the Central Statistics Office (CSO) which was sorted by electoral divisions.

There were a few town names in the trees' dataset that didn't fit in with any electoral district: Saint Margaret's, Lanesborough, Ballboughal, and Airways Industrial, so I removed these towns from the dataset.

As I was also looking at house prices as a comparative statistic, I removed towns that didn't have separate house price values: Corduff, Seatown and Hartstown.

The tree dataset file from Fingal County Council was called "**data_smart_dublin_dataset_trees – 36000.csv**"

	Tree_ID	Address	Town	Tree_Species	Species_Desc	Common_Name
1	27092	Clonard Court, Balb	Balbriggan	ACSA	Acer saccharinur	Silver Maple
2	29144	Ridgewood Avenue,	Swords South	TICO	Tilia cordata	Small-Leafed Lime
3	29160	Ridgewood Avenue,	Swords South	TICO	Tilia cordata	Small-Leafed Lime
4	29400	Ridgewood Avenue,	Swords South	TICO	Tilia cordata	Small-Leafed Lime
5	246560	Lanesborough Grov	Santry	TICO	Tilia cordata	Small-Leafed Lime
6	18012	Charlestown Drive,	Santry	TICO	Tilia cordata	Small-Leafed Lime
7	18016	Charlestown Drive,	Santry	TICO	Tilia cordata	Small-Leafed Lime

Figure 1- Imported Trees Dataset

Population Demographics

The Central Statistics Office website provides details of the population (Col **DV** with the heading "**T1_2T**" of all electoral divisions in a csv file: AllThemesTablesED.csv (CSO, 2020). I collated the information from the Census Office with information from the MyHome website to create an Excel sheet of electoral division demographic information (MyHome.ie, 2020).

To present demographic information effectively in ggplot charts, it was necessary to amalgamate columns. I did this with population ages, house ages, house type and house prices. This meant I could show this information as legends within ggplot graphs. I used the gather() and spread() functions to achieve this result.

The relevant population demographic file I loaded into R was called "**myHomeIEDemographics.csv**". It had nearly 400k records and was approx.: 56 Mb.

Town	Pop(Town	Popo(ED)	ED_unless	OnePersoı	CoupleWi	CoupleWi	SinglePare	OtherFam	RA<16	RA16-34	RA35-49	RA50-64	RA65+
Balbriggar	3619	24577	24577	20	17	38	16	9	31	29	24	11	5
Baldoyle	5016	14167	14167	20	23	35	11	11	20	28	23	16	13
Balrothery	720	NA	720	18	60	40	17	9	34	31	24	8	3
Blanchard	7491	104539	104539	20	22	31	9	18	20	34	21	12	13
Castlekno	8059	35002	35002	16	27	37	7	13	14	37	16	22	11
Clonee	9500	NA	9500	16	17	47	9	11	26	28	24	14	8
Clonsilla	7136	NA	7136	14	13	38	18	17	26	32	22	15	5
Donabate	162	9516	9516	19	18	45	10	8	30	22	30	12	6
Howth	5577	14565	14565	22	29	34	8	7	18	20	20	20	22

Figure 2 - Imported Population Dataset

House Prices

House prices was chosen as a data characteristic for its capacity to indicate a neighbourhood's prosperity. House prices can also be a good indicator of a neighbourhood's income level. If a city council is seeking to benefit those that are least well-off, then looking at house prices is as good an indicator of a neighbourhood's social need as any other.

The property prices dataset was obtained from the government's Property Services Regulatory Authority website (Property Services Regulatory Authority, 2020). This dataset covered all sales of property in Ireland from 2010 to present and was called "**PPR_ALL.csv**"

Date of Sale (dd/	Address	Postal Code	County	Price (€)	Not Full M	VAT Exclu:	Descriptio	Property Size Description
01/01/2010	5 Braemor Drive, Churchtown, Co.Dublin		Dublin	€343,000.00	No	No	Second-Hand Dwelling house /Apartı	
03/01/2010	134 Ashewood Walk, Summerhill Lane, Portlaoise		Laois	€185,000.00	No	Yes	New Dwel	greater than or equal to 38
04/01/2010	1 Meadow Avenue, Dundrum, Dublin 14		Dublin	€438,500.00	No	No	Second-Hand Dwelling house /Apartı	
04/01/2010	1 The Haven, Mornington		Meath	€400,000.00	No	No	Second-Hand Dwelling house /Apartı	
04/01/2010	11 Melville Heights, Kilkenny		Kilkenny	€160,000.00	No	No	Second-Hand Dwelling house /Apartı	
04/01/2010	12 Sallymount Avenue, Ranelagh		Dublin	€425,000.00	No	No	Second-Hand Dwelling house /Apartı	
04/01/2010	13 Oakleigh Wood, Dooradoyle, Limerick		Limerick	€172,500.00	No	No	Second-Hand Dwelling house /Apartı	

Figure 3 - Imported Property Prices Dataset

The property prices data required data cleansing, so I performed the following cleaning steps:

- Removed records that had not reached their recommended sale price.
- Removed superfluous columns
- Formatted the pricing column to numeric
- Used the as.date() function to format the sale date value to a date format so I could exclude properties sold before January 1st 2015 in the mean price calculation

The dataset did not show town names separately from the address so it was necessary to create a town names column to correspond with the CSO and tree datasets. Although the property price dataset included data for the whole of Ireland, I only needed property prices for the neighbourhoods in Fingal from 2015. I figured property prices before 2015 might adversely skew house price calculations. To limit the house prices to Fingal, I created a vector with the Fingal town names. I used the dplyr mutate(), sub() and paste() functions to detect towns in the Address column and used their combination to create a Town column.

```
sub(paste0("^(?:.*(", paste(townsP,collapse = "|"), ").*|.*)$"), "\\1", Address))
```

Of the 400k property price details I had in the original data source going back to 2010, 85k observations related to Dublin and only 4k related to Fingal.

Once I narrowed down the properties' dataset to properties in Dublin from 2015, I noticed that were some properties in Co Meath were part of Fingal. They were in the town of Clonee, so I used the rbind() function to includes sales in Clonee, Co Meath.

Insights

The aim of the project was to capture a profile of neighbourhoods in the form of demographics. By highlighting societal factors with tree coverage, it was possible to present a case for planting more trees to target particular social demographic in an area. Currently, city councils keep an inventory of trees for tree maintenance purposes and not necessarily for their ability to promote wellbeing in a neighbourhood. Of course, tree departments are aware of how trees benefit a community but city councils are not actively promoting the societal benefits of their tree planting policies. In addition to the wider community wellbeing benefit of trees, trees counteract CO_2 emissions (Arbor Day Foundation, 2020) and there are many more environmental and mental health benefit of trees.

Insight 1

(Line 88 rstudio)

Taking a look at tree coverage in Figure 4r, we can see that the spread of trees across neighbourhoods is quite uneven. Blanchardstown, Castleknock, Malahide and Swords have the most trees, but these neighbourhoods have bigger populations and are larger in physical size.

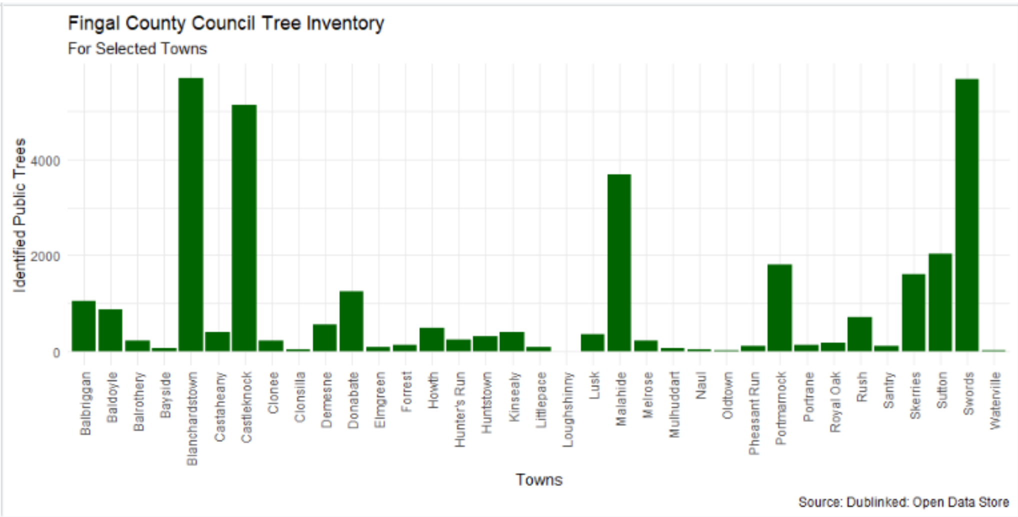

Figure 4 – Fingal Trees

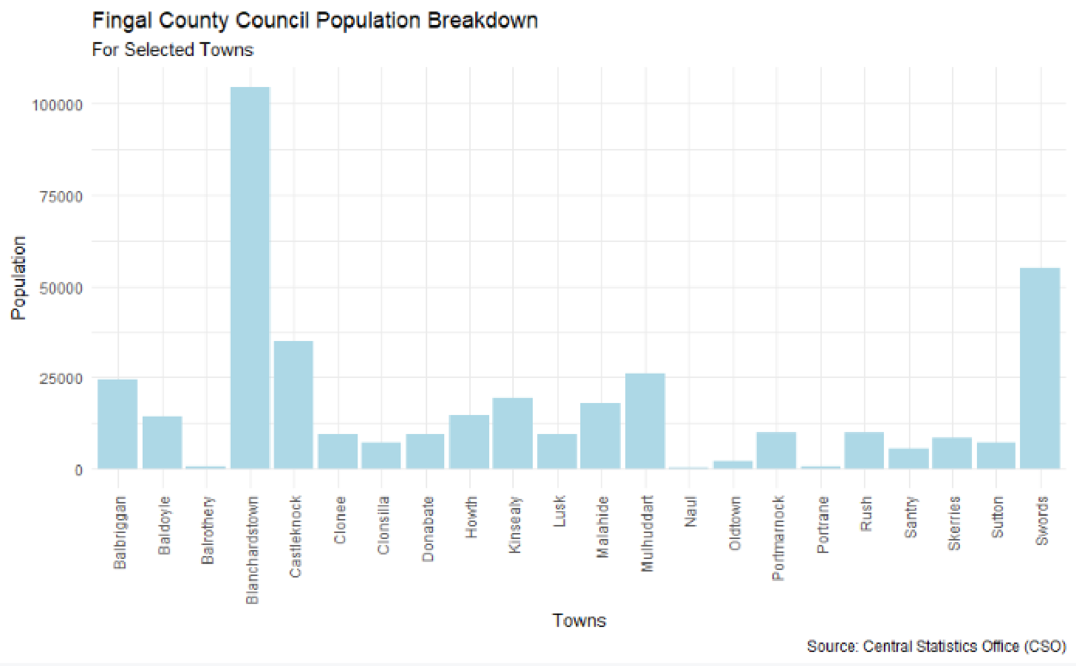

Figure 5 – Fingal Population

The population of Blanchardstown, Castleknock and Swords are also high so we can see that there is a correlation between these population and tree coverage in these neighbourhoods.

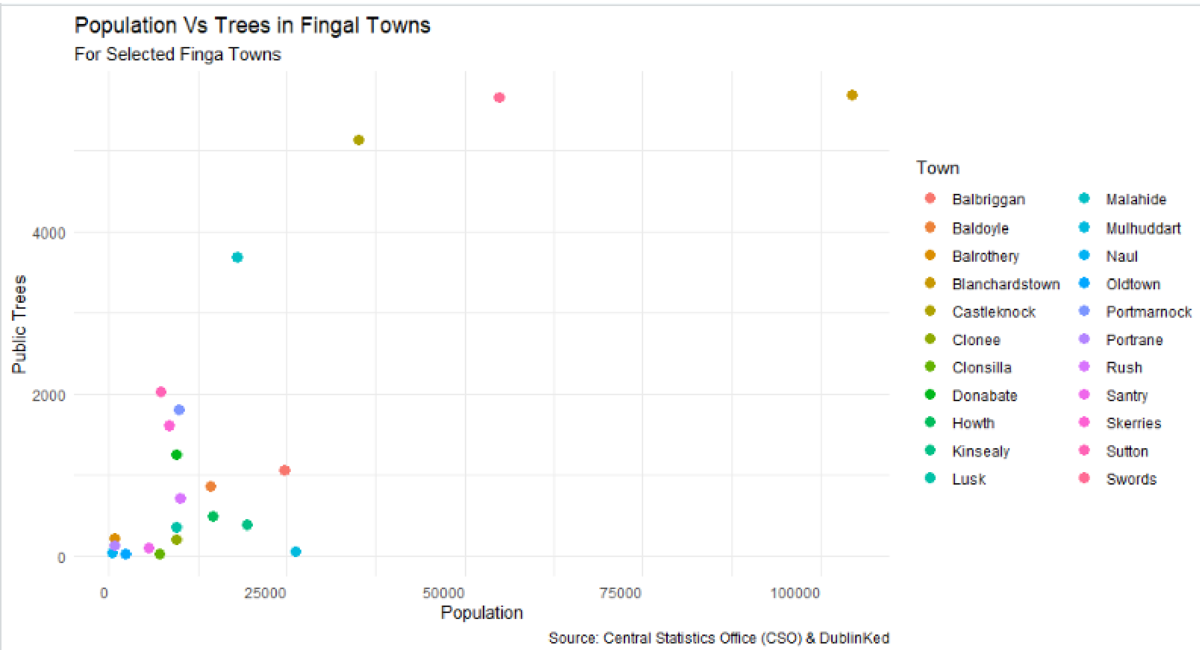

Figure 6 - Insight 1a - Relationship between Pop and Trees

```
> lm(dublinPop2$Pop ~ dublinPop2$TotalTrees)

Call:
lm(formula = dublinPop2$Pop ~ dublinPop2$TotalTrees)

Coefficients:
          (Intercept)  dublinPop2$TotalTrees
             3878.091                  9.711

> cor(dublinPop2$Pop, dublinPop2$TotalTrees)
[1] 0.785019
```

Figure 7 - Insight 1b - Correlation between Population and Trees

The correlation is .785019, so is a strong correlation. Looking at the linear regression model, for example you would expect 2,574 trees for a town with a population of 25,000 people (25000/9.711). If discussing with the council about their tree planting policy, it would be good to look at this correlation and compare this correlation with tree planting around the rest of the world and whether the trees in a particular town is less or more than the Fingal average.

Insight 2

(Line261 & 330 rstudio)

During the investigation of tree planting by Fingal council, I looked at house prices by way of gauging the socio-economic profile of the town residents. Taking the property sales between 2015 and 2019 I derived a mean property price for house prices in Fingal.

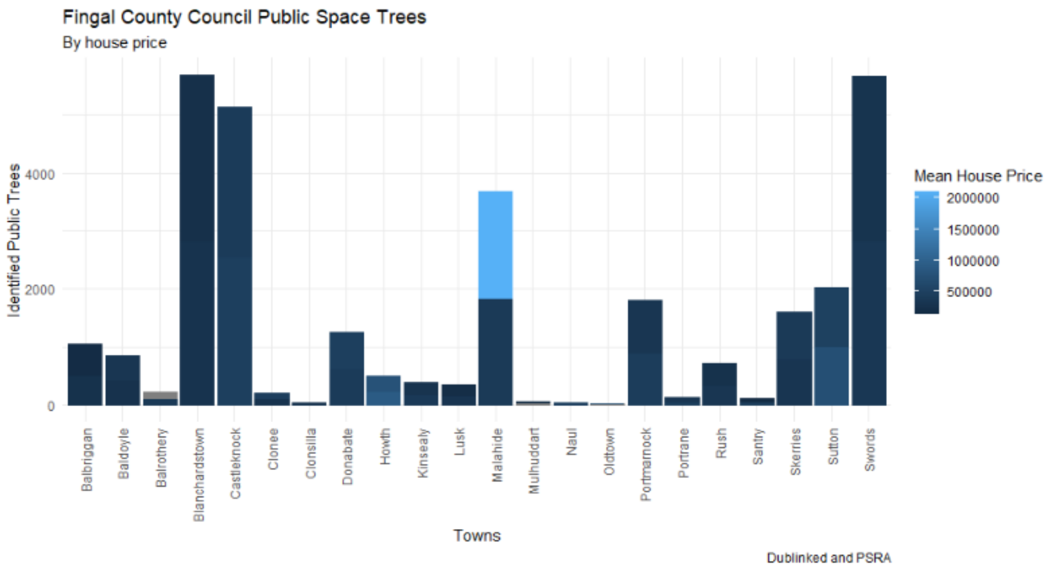

Figure 8 - Trees Vs House Prices

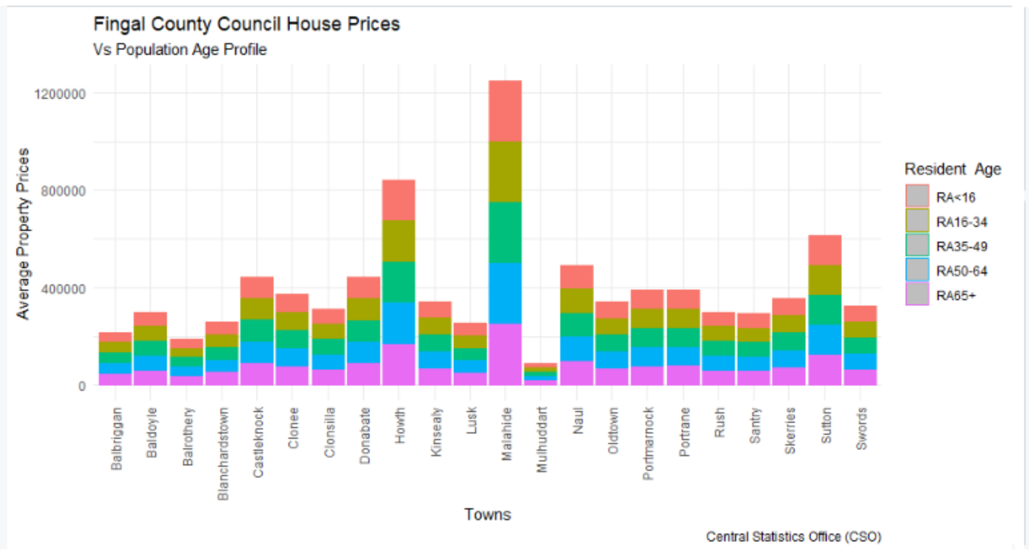

Figure 9 – - Property Prices with following age profile of residents

When calculating the mean prices in Figure 8 and 9 graphs, I confined the property prices dataset to the years 2015 to 2019 so property prices were not unnecessarily skewed by low prices during the period 2010 to 2014.

However, I noticed when looking at mean prices during the years between 2015 and 2019 that the mean price in 2016 and 2017 were not that different and I wondered if this difference was significantly different in statistical terms

```
> t.test(propertiesMeanPrice2016$Price, propertiesMeanPrice2017$Price)

        Welch Two Sample t-test

data:  propertiesMeanPrice2016$Price and propertiesMeanPrice2017$Price
t = -1.5194, df = 79726, p-value = 0.1287
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -39521.587   5004.752
sample estimates:
mean of x mean of y
 470862.7  488121.1
```

Figure 10 - Insight 2 - Property Prices 2016/17 t.test

I ran a t-test to test this hypothesis, see Figure 9. The p-value associated with the test was 0.1287, so I was not able to reject the null hypothesis H_o that there is no difference between the true mean of the two groups since the p-value is greater than the non-critical significance alpha level of .05. Based on this, we can conclude that there is not enough evidence of a difference between the mean of the two groups to reject the null hypothesis that they are the same so the difference is mean prices is statistically significant.

Insight 3
(Line 338 rstudio)

Here I looked at house prices for the Fingal area over the years I used to calculate the mean house prices. I saw that there were some outliers.

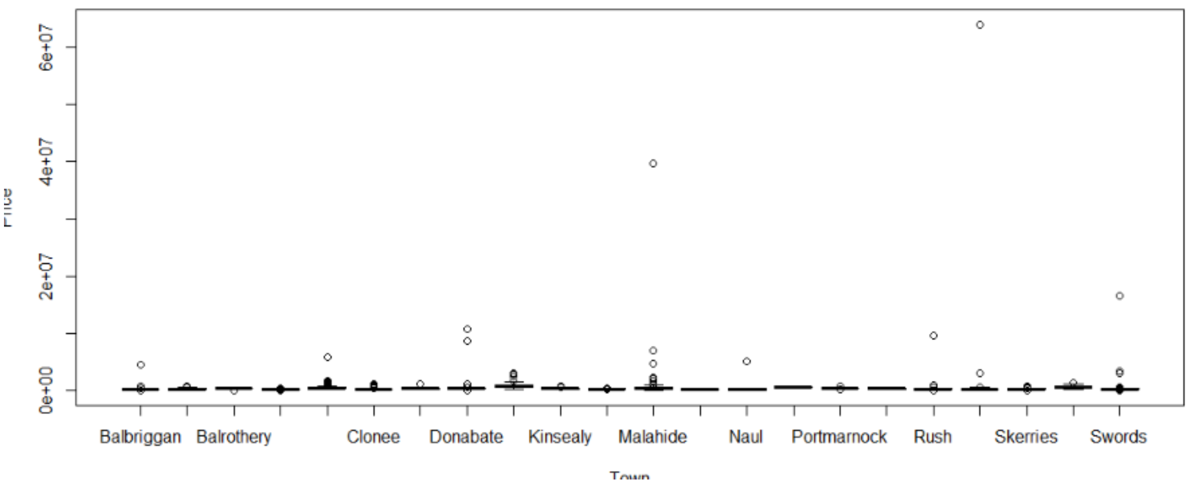

Figure 11 – Insight 3 - Boxplot of House Prices by Town

We can see that there were some extreme outliers in the period 2015 to 2019 for houses in the Malahide and Santry towns. Santry is located between Rush and Skerries in the above chart. These outliers would have skewed the mean calculations for these districts. This is an argument for excluding these outliers from the dataset.

Insight 4

(Line 347 rstudio)

Correlation of second-hand and new property prices in Fingal.

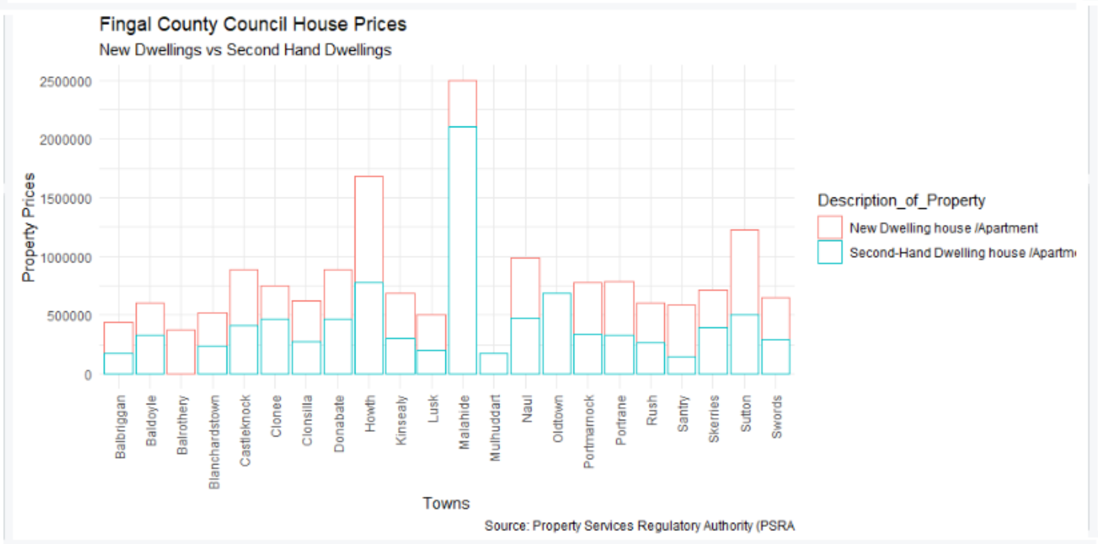

Figure 12 - Property Prices for Old and New Homes

New house property prices are invariably higher than second-hand property prices.

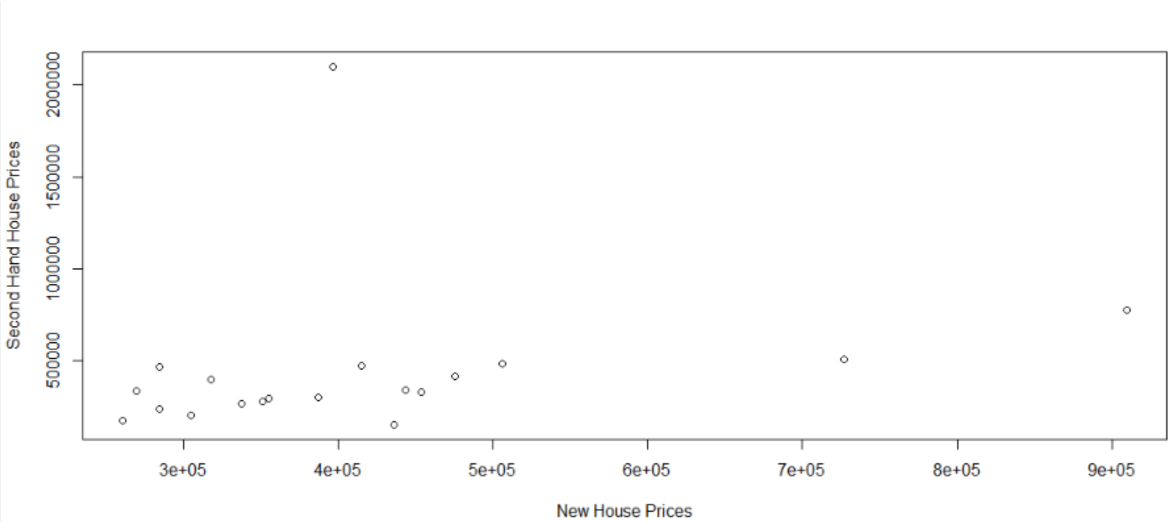

Figure 13 - Plot of Old Vs New House Prices

```
> lm(dublinPop2$`New Dwelling house /Apartment` ~ dublinPop2$`Second-Hand Dwelling house /Apartment`)

Call:
lm(formula = dublinPop2$`New Dwelling house /Apartment` ~ dublinPop2$`Second-Hand Dwelling house /Apartment`)

Coefficients:
                                  (Intercept)  dublinPop2$`Second-Hand Dwelling house /Apartment`
                                    3.768e+05                                           8.902e-02
```

Figure 14 - Insight 4

Conclusion

The main findings of the project that population and property prices are related and property prices are not constant but are changing significantly over the years. Tree planting is related to population size so this is a positive but the question to ask is whether this correlation is enough and if the correlation could be improved. There were some anomalies in house prices with outliers and this needs to be investigated and those house prices need to be removed to prevent these outliers from distorting house value averages for an area.

Challenges

I spent a considerable amount of time sourcing raw data. The effort allowed me to come up with insights that are not normally available to the public. This was worth the effort as it is much more rewarding exploring commonly unknown insights.

Bibliography

Arbor Day Foundation, 2020. *Trees Help Fight Climate Change.* [Online]
Available at: https://www.arborday.org/trees/climatechange/
[Accessed 27 April 2020].
Coillte, 2020. *HEALTH AND WELL BEING.* [Online]
Available at: https://www.coillte.ie/our-forests/public-goods/health-and-well-being/
[Accessed 26 April 2020].
CSO, 2020. *Census 2011 Small Area Population Statistics (SAPS).* [Online]
Available at: Census 2011 Small Area Population Statistics (SAPS)
[Accessed 20 April 2020].
Lynn, S., 2018. *The Irish Property Price Register – Geocoded to Small Areas.* [Online]
Available at: https://www.shanelynn.ie/the-irish-property-price-register-geocoded-to-small-areas/
[Accessed 20 April 2020].
MyHome.ie, 2020. *Ireland Neighbourhood Guide.* [Online]
Available at: https://www.myhome.ie/neighbourhood-guide
[Accessed 20 April 2020].
Property Services Regulatory Authority, 2020. *Property Services Regulatory Authority.* [Online]
Available at:
https://www.propertypriceregister.ie/website/npsra/pprweb.nsf/PPRDownloads?OpenForm&File=PPR-2019-08-Dublin.csv&County=Dublin&Year=2019&Month=08
[Accessed 20 April 2020].
Smart Dublin - Dublinked, 2020. *Trees.CSV.* [Online]
Available at: https://data.smartdublin.ie/dataset/trees/resource/47d28489-44e5-4eee-93a8-a31e110f6cce?view_id=153603c7-be83-403d-b2a0-fdb14c1a4bd7&filters=
[Accessed 16 April 2020].
The Royal Parks, 2020. *Why are trees so important?.* [Online]
Available at: https://www.royalparks.org.uk/parks/the-regents-park/things-to-see-and-do/gardens-and-landscapes/tree-map/why-trees-are-important
[Accessed 26 April 2020].

Appendix – R code

```
## Mairead Manifold x19207034 Prog for Big Data - CA - Project - April 2020
## Lecturer: Dr Eugene O'Loughlin

library(readr)
library(dplyr)
library(tidyr)
library(stringr)
library(ggplot2)
theme_set(
  theme_minimal() +
    theme(legend.position = "right"))


## DUBLIN Trees
dublinTrees <- read_csv("data_smart_dublin_dataset_trees - 36000.csv")  ## 35,925 obs

dublinTrees$Town <- as.factor(dublinTrees$Town)
unique(dublinTrees$Town) ## identify towns in Fingal and created a vector with these town names.
dublinTrees[is.na(dublinTrees$Town),] ## 845 are NAs

towns <- c("Airways Industrial", "Balbriggan", "Baldoyle", "Bayside", "Corduff", "Castlehaney", "Balgriffin", "Ballyboughal",
"Balrothery", "Blanchardstown", "Castaheany", "Castleknock", "Carpenterstown", "Clonee", "Clonsilla", "Demesne",
"Donabate", "Elmgreen", "Forrest", "Garristown", "Hartstown", "Howth", "Huntstown", "Hunter's Run",  "Kinsealy",
"Lanesborough", "Littlepace", "Loughshinny", "Lusk", "Malahide", "Melrose", "Mulhuddart", "Naul", "Oldtown","Pheasant
Run", "Portmarnock", "Portrane", "Royal Oak", "Rush", "Santry", "Saint Margaret", "Seatown", "Skerries", "Sutton", "Swords",
"Waterville")


## Several enteries don't have the town listed. Using following function to locate towns in the address field
## and update the missing town details in the town field. There are 845 Town NAs where an address is give. Running following
function
## reducted the number of NAs from 845 to 67
dublinTrees <- dublinTrees %>% mutate(Town =  sub(paste0("^(?:.*(", paste(towns,collapse = "|"), ").*|.*)$"), "\\1",
Address))


dublinTrees[is.na(dublinTrees$Town),] ## no of NAs reduced to 67


dublinTrees <- dublinTrees[!is.na(dublinTrees$Town),] ## 35,857 observations remaining

## Merge some towns to their nearest neighbour town district in instances where I don't have either
## corresponding population or property price data

dublinTrees[dublinTrees$Town == "Bayside", "Town",] <- "Baldoyle"

dublinTrees[dublinTrees$Town == "Castaheany" | dublinTrees$Town == "Hunter's Run" | dublinTrees$Town ==
"Huntstown" | dublinTrees$Town == "Pheasant Run", "Town"] <- "Mulhuddart"

dublinTrees[dublinTrees$Town == "Demesne" | dublinTrees$Town == "Forrest" | dublinTrees$Town == "Melrose",
"Town"] <- "Swords"

dublinTrees[dublinTrees$Town == "Loughshinny", "Town"] <- "Rush"

dublinTrees[dublinTrees$Town == "Elmgreen" | dublinTrees$Town == "Littlepace" | dublinTrees$Town == "Waterville",
"Town"] <- "Blanchardstown"

dublinTrees[dublinTrees$Town == "Royal Oak", "Town"] <- "Santry"

unique(dublinTrees$Town) ## checking what towns I have narrowed down my selection to = 43



## remove Airways Industrial, Ballboughal, Lanesborough and Saint Margaret's as there are not suitable for
## incorporation in other towns and I don't have comparative data.
```

```
dublinTrees <- dublinTrees %>%
  filter(Town != "Airways Industrial" & Town != "Ballyboughal" & Town != "Lanesborough" & Town != "Saint Margaret")

## similar issue - removing towns where I don't have any property price details: Corduff, Seatown and Hartstown

dublinTrees <- dublinTrees %>% ## removing unwanted towns
  filter(Town != "Corduff" & Town != "Seatown" & Town != "Hartstown" & Town != "" )


dublinTrees %>%  ## 1,699  of the Common Name have NAs
  filter(is.na(dublinTrees$Common_Name))


sum(!is.na(dublinTrees$Common_Name))## 31,969 tree names can be identified


dublinTrees %>%  ## 9,846 of 28,862
  filter(Age_Desc == "Semi-Mature")

dublinTrees %>%  ## 34 of 28,862
  filter(Age_Desc == "Over-Mature")

dublinTrees %>%  ## 5,988 of 28,862
  filter(Age_Desc == "Young")

dublinTrees %>%  ## 1,184 of 28,862
  filter(Age_Desc == "Newly Planted")

dublinTrees %>%  ## 19 of 28,862
  filter(Age_Desc == "Veteran")

  ## INSIGHT 1 - Trees per neighbourhood in Fingal

treesByTown <-
  dublinTrees %>%
  group_by(Town) %>%
  summarize(TotalTrees = sum(!is.na(Tree_ID))) ## totals to 32,971

ggplot(treesByTown, aes(x = Town, y = TotalTrees )) +
   geom_col(fill = "dark green") +
  labs(x = "Towns", y = "Identified Public Trees", title = "Fingal County Council Tree Inventory",
       subtitle = "For Selected Towns",
       caption = "Source: Dublinked: Open Data Store") +
       theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))


ggplot(dublinPop, aes(x = Town, y = Pop )) +
  geom_col(fill = "light blue") +
  labs(x = "Towns", y = "Population", title = "Fingal County Council Population Breakdown",
       subtitle = "For Selected Towns",
       caption = "Source: Central Statistics Office (CSO)") +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))



ggplot(dublinPop2, aes(x = Pop, y = TotalTrees, color = Town )) +
  geom_point(size = 3) +
  labs(x = "Population", y = "Public Trees", title = "Population Vs Trees in Fingal Towns ",
       subtitle = "For Selected Finga Towns",
       caption = "Source: Central Statistics Office (CSO) & DublinKed") +
  theme(axis.text.x=element_text(angle=0,hjust=1,vjust=0.5))

lm(dublinPop2$Pop ~ dublinPop2$TotalTrees)

cor(dublinPop2$Pop, dublinPop2$TotalTrees)
```

```
## END OF INSIGHT 1


##PROPERTIES - House prices DATA CLEANSING

properties <- read_csv("PPR-ALL.csv", col_names = c("Date_of_Sale", "Address", "Postal_Code", "County",
"Price","Not_Full_Market_Price", "Vat_Exclusive", "Description_of_Property", "Property_Size_Description"))


## keep properties that reached full value only as they may skew the results
properties %>%
  filter(Not_Full_Market_Price == "Yes") # = 20,396 rows, so removing them

#remove 20,396 rows so now have 398,797
properties <- properties %>%
  filter(Not_Full_Market_Price == "No") # = Removing 20,396 rows and left with 398,787

# drop the columns: "Postal_Code"(3), Not_Full_Market_Price" (6), "Vat_Exclusinve" (7)
properties <- properties[-c(3,6,7)] ## 6 variables remaining


unique(properties$Description_of_Property) #identify Irish language dwelling descriptions

## subset properties to only English language property descriptions - removed 31 so down to 398,766
properties <- properties %>%
  filter(Description_of_Property == "Second-Hand Dwelling house /Apartment" | Description_of_Property =="New Dwelling
house /Apartment")

## subset properties to only those after 2015 where dates are in the format day/month/year and in character class
unique(properties$Date_of_Sale)

#changed date field to date format.
properties$Date_of_Sale <- as.Date(properties$Date_of_Sale, format = "%d/%m/%Y")
unique(properties$Date_of_Sale)

## reduced list down to 267,123
properties <- properties %>%
  filter(Date_of_Sale > "2015-01-01") %>%
  arrange(Date_of_Sale)



unique(properties$Price)

##using stringr package, remove the pence from price.

properties$Price <- str_sub(properties$Price, end=-3)

## alternative data cleansing - properties$Price <- gsub(".{3}$", "", properties$Price)

## alternative data cleansing - pproperties$Price = substr(properties$Price, 1, nchar(properties$Price)-3)


##removing all other non numerical characters from price such as x80, commas and full stop.

properties$Price <- gsub("[^A-Za-z0-9]", "", properties$Price)
unique(properties$Price)
# converting prices variable to numeric and updating other variable formats
properties$Price <- as.numeric(properties$Price)
properties$Description_of_Property <- as.factor(properties$Description_of_Property)
properties$Property_Size_Description <- as.factor(properties$Property_Size_Description )

townsP <- c("Balbriggan", "Baldoyle", "Balrothery", "Blanchardstown", "Castleknock",  "Clonee", "Clonsilla", "Corduff",
"Donabate", "Hartstown", "Howth", "Kinsealy", "Lusk", "Malahide", "Mulhuddart", "Naul", "Oldtown","Portmarnock",
"Portrane", "Rush", "Santry", "Seatown", "Skerries", "Sutton", "Swords")

## Find town in Address and use it to create a Town column which does not exist
properties <- properties %>% mutate(Town =  sub(paste0("^(?:.*(", paste(townsP,collapse = "|"), ").*|.*)$"), "\\1", Address))
```

```
unique(properties$Town)


properties %>%  ## 85,681 of these are in Dublin county
  select(Date_of_Sale, Address, County, Price, Town) %>%
  filter(County == "Dublin")

properties %>%  ## 181,441 of these are outside Dublin
  select(Date_of_Sale, Address, County, Price, Town) %>%
  filter(County != "Dublin")


## convert "" in Town to NAs
properties$Town <- gsub("^$|^ $", NA, properties$Town)

properties %>% #4,673
  filter(!is.na(Town)) ## only 4,673 relate to Fingal towns I am investigating

#checking no "" remaining
unique(properties$Town)

properties %>%  ## 4,543
  filter(Town %in% towns) ## 4,872 in total

properties %>%  ## 4,543
  filter(Town %in% towns, County == "Dublin")


## looking at the town distribution, only the Clonee in Co Meath should
## be included in my property prices list.
propertiesCloneeMeath <- properties %>%  ## 111
  filter(Town %in% towns, County == "Meath", Town == "Clonee")


propertiesMyTowns <- properties %>%  ## 4,543
  filter(Town %in% towns, County == "Dublin")

##combine the Clonee properties in Co Meath in my properties data set
## using the rbind() function.
propertiesMyTowns <- rbind(propertiesMyTowns, propertiesCloneeMeath) ## 5,175

propertiesMyTowns$Town <- as.factor(propertiesMyTowns$Town)
dublinTrees$Town <- as.factor(dublinTrees$Town)

unique(propertiesMyTowns$Town)
unique(dublinTrees$Town)


## DUBLIN POPULATION Dataset

dublinPop <- read_csv("myHoneIEDemographics.csv")  ## 35,925 obs
unique(dublinPop$Town)


## To include property prices in the dublinPopulation table, I used the spread() function
housePricesByPropertyType <- spread(propertiesMyTownsMeanPriceByPrDe, Description_of_Property, meanPrice)

adddf_prices <- merge(dublinPop, housePricesByPropertyType, by = "Town")
adddf_trees <- merge(adddf_prices, treesByTown, by = "Town")

## to facilitate ggplot inclusion of a third variable with legends, I used the gather() function to group
## related columns together and make it possible to include the columns in a ggplot graph

dublinPop2 <- gather(adddf_trees, FamilyType, PerFamType, 5:9)
dublinPop3 <- gather(dublinPop2, ResAge, PerResAge, 5:9)
dublinPop4 <- gather(dublinPop3, HouseAge, PerHouseAge, 5:10)
dublinPop5 <- gather(dublinPop4, HouseType, PerHouseType, 5:7)
```

```
dublinPop6 <- gather(dublinPop5, HouseOld_New, MeanPriceHouse, 5:6)



## INSIGHT 2 - Mean Property prices across Ireland in 2015 Vs 2019 - using t-test to see if they
## are the same
## t.test mean property price 2015 versu mean property price 2019

## gives you mean house price by Town district
propertiesMyTownsMeanPriceByPrDe <-
  propertiesMyTowns %>%
  group_by(Town, Description_of_Property) %>%
  summarize(meanPrice = mean(Price))


## Trees by age profile of residents
ggplot(dublinPop6, aes(x = Town, y = TotalTrees/900, color = MeanPriceHouse )) +
  geom_col() +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  labs(x = "Towns", y = "Identified Public Trees", title = "Fingal County Council Public Space Trees",
       subtitle = "By house price",
       caption = "Dublinked and PSRA", color = "Mean House Price")



ggplot(dublinPop6, aes(x = Town, y = MeanPriceHouse/900, color = ResAge )) +
  geom_col(fill = "grey") +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))  +
  labs(x = "Towns", y = "Average Property Prices", title = "Fingal County Council House Prices",
       subtitle = "Vs Population Age Profile",
       caption = "Central Statistics Office (CSO)", color = "Resident  Age")



propertiesMeanPrice2015 <- properties %>% ## 72k observations
  filter(Date_of_Sale >= "2015-01-01" & Date_of_Sale >= "2015-12-01", County == "Dublin") %>%
  arrange(Date_of_Sale) %>%
  mutate(meanPropertyPriceDublin_2015 = mean(Price)) ## don't need this

propertiesMeanPrice2015 <-propertiesMeanPrice2015 [-c(1:3,5:8)]

propertiesMeanPrice2016 <- properties %>% ## 57k observations
  filter(Date_of_Sale >= "2016-01-01" & Date_of_Sale >= "2016-12-01", County == "Dublin") %>%
  arrange(Date_of_Sale) %>%
  mutate(meanPropertyPriceDublin_2016 = mean(Price)) ## don't need this

propertiesMeanPrice2016 <-propertiesMeanPrice2016 [-c(1:3,5:8)]

propertiesMeanPrice2017 <- properties %>% ## 41k observations
  filter(Date_of_Sale >= "2017-01-01" & Date_of_Sale >= "2017-12-01", County == "Dublin") %>%
  arrange(Date_of_Sale) %>%
  mutate(meanPropertyPriceDublin_2017 = mean(Price))  ## don't need this

propertiesMeanPrice2017 <-propertiesMeanPrice2017  [-c(1:3,5:8)]

propertiesMeanPrice2018 <- properties %>% ## 23k observations
  filter(Date_of_Sale >= "2018-01-01" & Date_of_Sale >= "2018-12-01", County == "Dublin") %>%
  arrange(Date_of_Sale) %>%
  mutate(meanPropertyPriceDublin_2018 = mean(Price)) ## don't need this

propertiesMeanPrice2018 <-propertiesMeanPrice2018  [-c(1:3,5:8)]

propertiesMeanPrice2019 <- properties %>% ## 5k observations
  filter(Date_of_Sale >= "2019-01-01" & Date_of_Sale >= "2019-12-01", County == "Dublin") %>%
```

```
  arrange(Date_of_Sale) %>%
  mutate(meanPropertyPriceDublin_2019 = mean(Price)) ## don't need this

propertiesMeanPrice2019 <-propertiesMeanPrice2019 [-c(1:3,5:8)]


##T.TEST INSIGHT 2

t.test(propertiesMeanPrice2016$Price, propertiesMeanPrice2017$Price)


## END OF INSIGHT 2


## INSIGHT 3

## Checking the presence of outliers in property prices I used to calculate mean property prices for the various
## towns under review in Fingal.


boxplot(Price ~ Town, data = propertiesMyTowns)


## INSIGHT 4


## gives you mean house price by Town district
propertiesMyTownsMeanPriceByPrDe <-
  propertiesMyTowns %>%
  group_by(Town, Description_of_Property) %>%
  summarize(meanPrice = mean(Price))

ggplot(propertiesMyTownsMeanPriceByPrDe, aes(x = Town, y = meanPrice, color = Description_of_Property )) +
  geom_col(fill = "white")  +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
  labs(x = "Towns", y = "Property Prices", title = "Fingal County Council House Prices",
       subtitle = "New Dwellings vs Second Hand Dwellings",
       caption = "Source: Property Services Regulatory Authority (PSRA)") +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))


lm(dublinPop2$`New Dwelling house /Apartment` ~ dublinPop2$`Second-Hand Dwelling house /Apartment`)

plot(dublinPop2$`New Dwelling house /Apartment`,dublinPop2$`Second-Hand Dwelling house /Apartment`, xlab = "New
House Prices", ylab = "Second Hand House Prices" )




#
```