Tutorial 2

Iz Leitch

Introduction

This is the paper about rentsafe toronto housing scores

Running Code

```
## Install libraries
#install.packages("janitor")
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
library(opendatatoronto)
library(dplyr)
library(scales)
```

```
Attaching package: 'scales'

The following object is masked from 'package:purrr':

    discard

The following object is masked from 'package:readr':

    col_factor
```

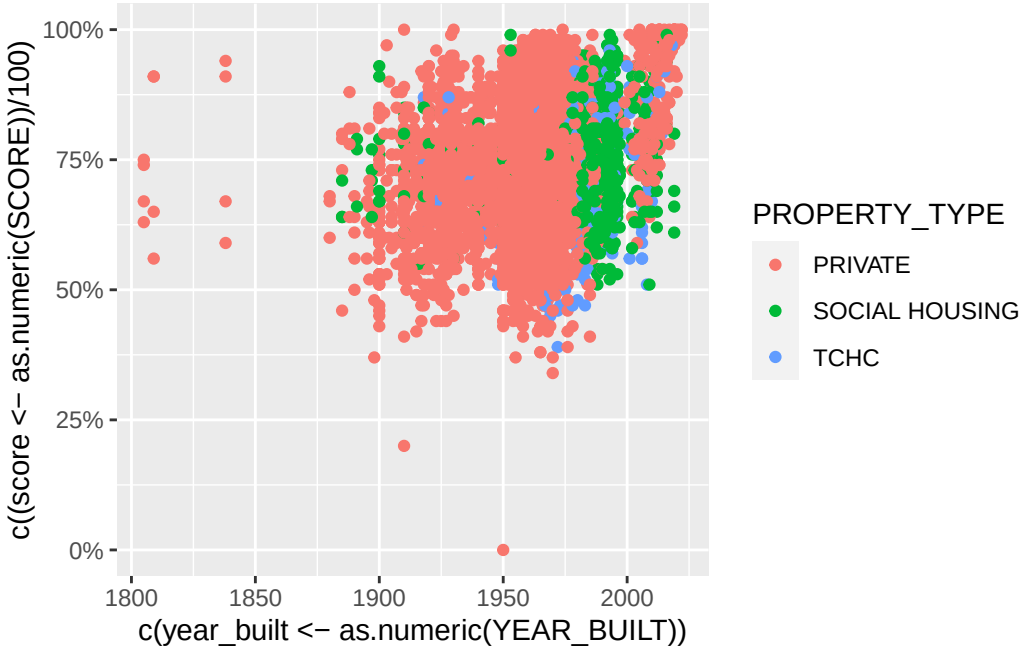
```
echo = FALSE

# get package
package <- show_package("4ef82789-e038-44ef-a478-a8f3590c3eb1")
package
```

```
# A tibble: 1 x 11
  title     id    topics civic~1 publi~2 excerpt datas~3 num_r~4 formats refre~5
  <chr>     <chr> <chr>  <chr>   <chr>   <chr>   <chr>     <int> <chr>   <chr>
1 Apartmen~ 4ef8~ Locat~ Afford~ Munici~ This d~ Table         4 XML,JS~ Daily
# ... with 1 more variable: last_refreshed <date>, and abbreviated variable
#   names 1: civic_issues, 2: publisher, 3: dataset_category, 4: num_resources,
#   5: refresh_rate
```

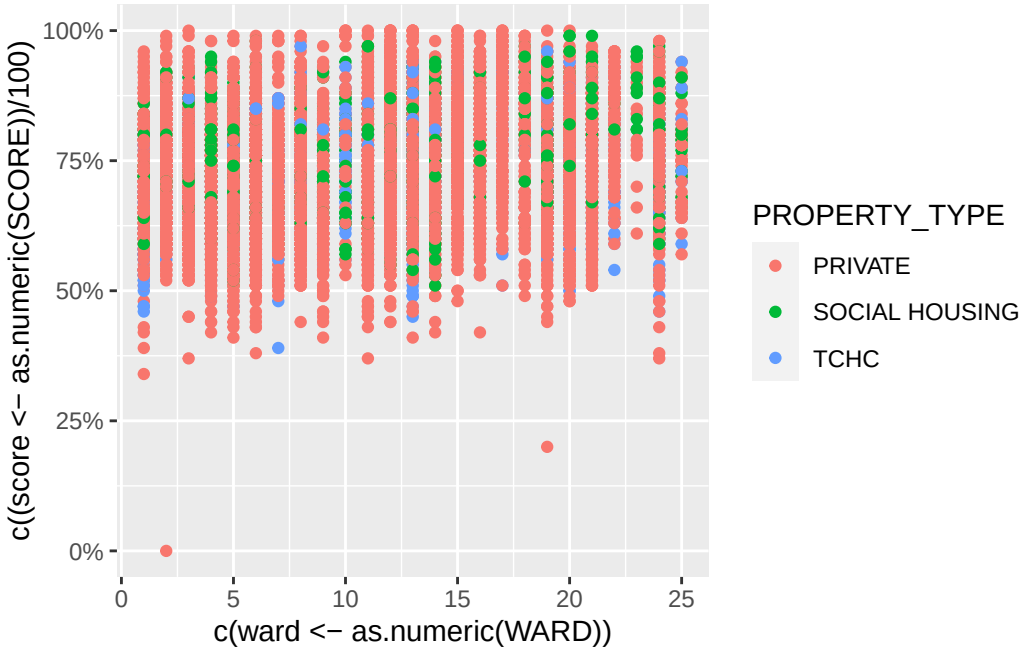
```
# get all resources for this package
resources <- list_package_resources("4ef82789-e038-44ef-a478-a8f3590c3eb1")

# identify datastore resources; by default, Toronto Open Data sets datastore resource form
datastore_resources <- filter(resources, tolower(format) %in% c('csv', 'geojson'))

# load the first datastore resource as a sample
data <- filter(datastore_resources, row_number()==1) %>% get_resource()
```

```
options(repr.plot.width = 5, repr.plot.height =2)

data|> ggplot(aes(x=c(year_built<-as.numeric(YEAR_BUILT)), y = c((score<-as.numeric(SCORE)
  scale_x_continuous(breaks = c(1800, 1850, 1900, 1950, 2000, 2050)) +
  geom_point()
```

```
Warning: Removed 48 rows containing missing values (`geom_point()`).
```

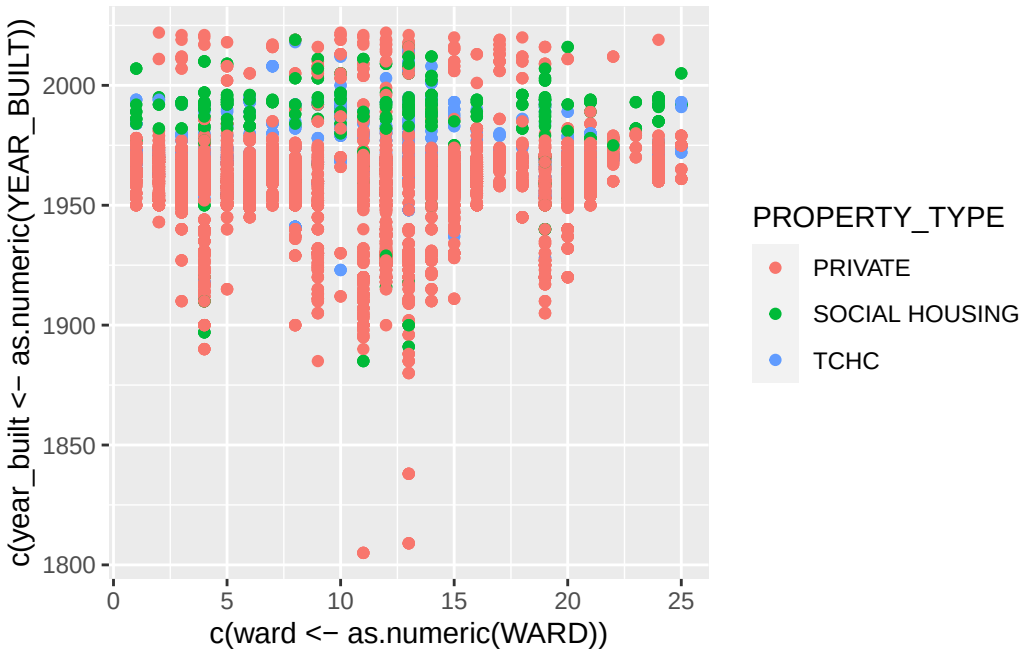



```
data|> ggplot(aes(x=c(ward<-as.numeric(WARD)), y = c((score<-as.numeric(SCORE))/100), colo
  geom_point()
```

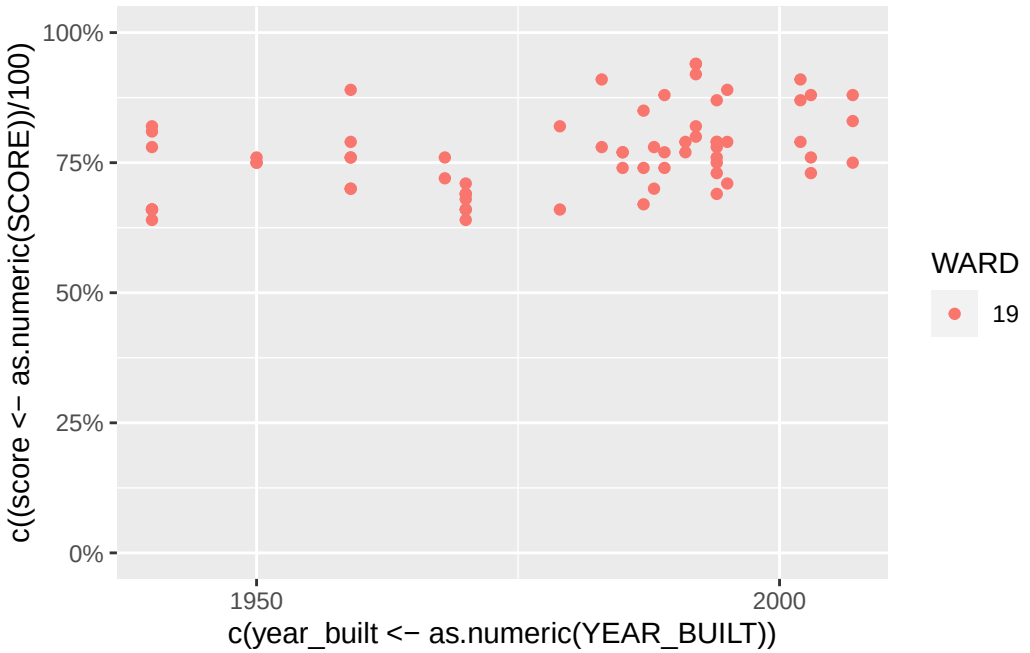



```
data|> ggplot(aes(x=c(ward<-as.numeric(WARD)), y = c(year_built<-as.numeric(YEAR_BUILT)),
  geom_point()
```

```
Warning: Removed 48 rows containing missing values (`geom_point()`).
```

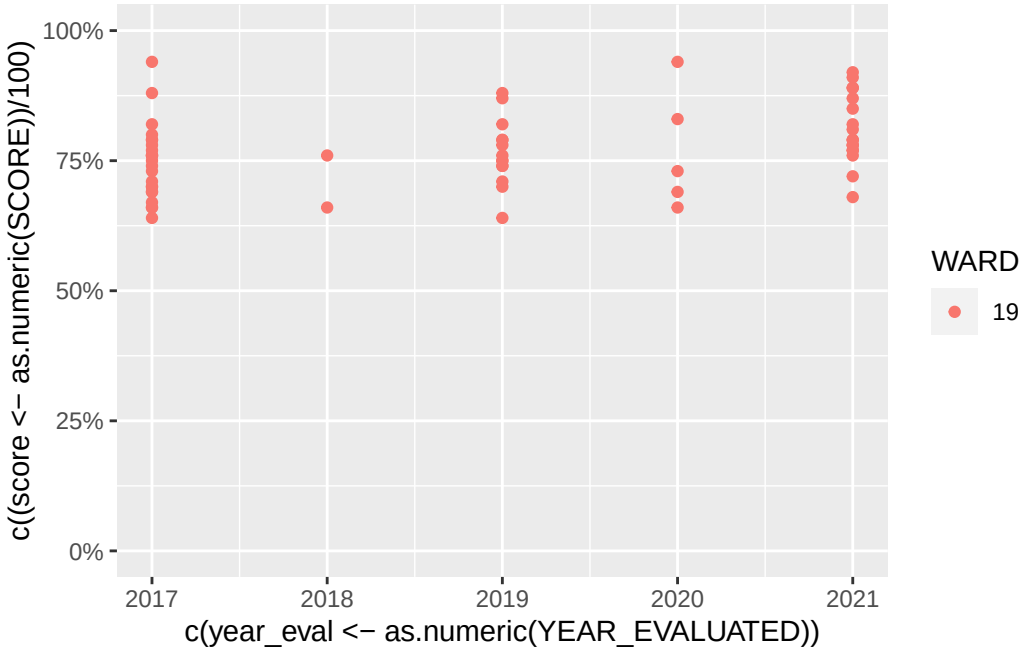



```
data|>
  filter(PROPERTY_TYPE == "SOCIAL HOUSING", WARD == "19") |>
  ggplot(aes(x=c(year_built<-as.numeric(YEAR_BUILT)), y = c((score<-as.numeric(SCORE))/100
  scale_x_continuous(breaks = c(1800, 1850, 1900, 1950, 2000, 2050)) +
  geom_point()
```



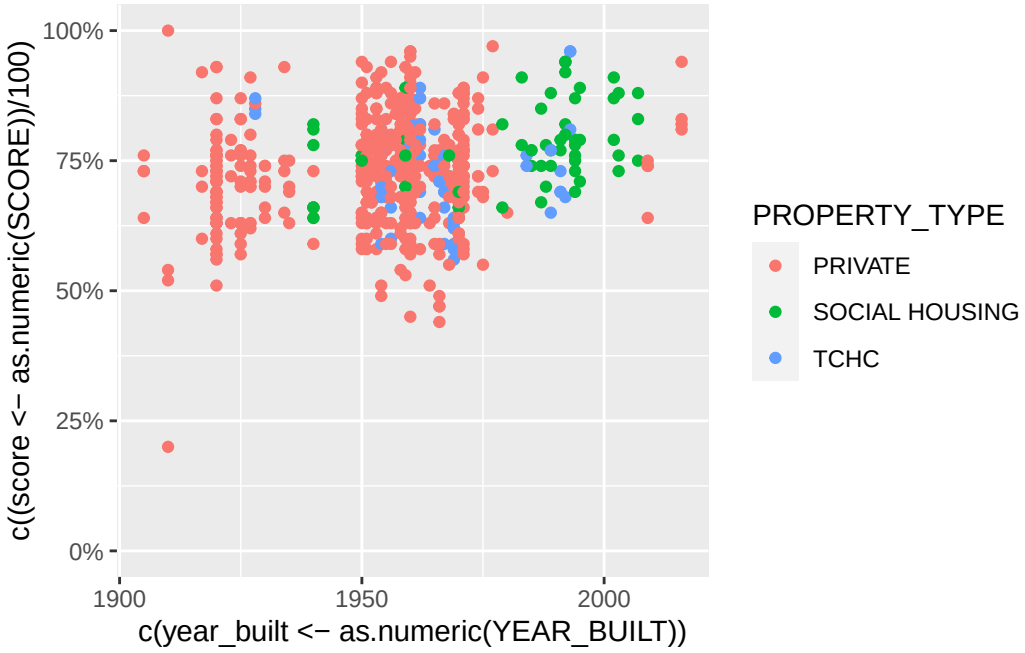

```
data|>
  filter(PROPERTY_TYPE == "SOCIAL HOUSING", WARD == "19") |>
  ggplot(aes(x=c(year_eval<-as.numeric(YEAR_EVALUATED)), y = c((score<-as.numeric(SCORE))/
  geom_point()
```

```
Warning: Removed 6 rows containing missing values (`geom_point()`).
```




```
data|>
  filter(WARD == "19") |>
  ggplot(aes(x=c(year_built<-as.numeric(YEAR_BUILT)), y = c((score<-as.numeric(SCORE))/100
  scale_x_continuous(breaks = c(1800, 1850, 1900, 1950, 2000, 2050)) +
  geom_point()
```

```
Warning: Removed 5 rows containing missing values (`geom_point()`).
```




```
data|>
  arrange(SCORE)
```

```
# A tibble: 11,651 x 40
     `_id` RSN     YEAR_~1 YEAR_~2 YEAR_~3 PROPE~4 WARD  WARDN~5 SITE_~6 CONFI~7
     <int> <chr>   <chr>   <chr>   <chr>   <chr>   <chr> <chr>   <chr>   <chr>
 1 2402709 4155384 2017    2017    1950    PRIVATE 02    Etobic~ 339 TH~ 5
 2 2394454 5186997 <NA>    <NA>    2019    PRIVATE 12    Toront~ 200 MA~ 6
 3 2394455 5118732 2022    <NA>    2021    PRIVATE 13    Toront~ 25 NIC~ 29
 4 2394460 5207679 2023    <NA>    2022    PRIVATE 12    Toront~ 215 LO~ 20
 5 2394461 5175953 2022    <NA>    2022    PRIVATE 10    Spadin~ 57 SPA~ 36
 6 2394872 4264094 2017    <NA>    2013    PRIVATE 13    Toront~ 132 BE~ 10
 7 2394997 4697372 2019    <NA>    2019    PRIVATE 17    Don Va~ 55 SMO~ 27
 8 2395236 4153423 2017    <NA>    2005    PRIVATE 13    Toront~ 167 CH~ 28
 9 2395301 4171374 2017    <NA>    2013    PRIVATE 10    Spadin~ 570 BA~ 29
10 2395431 4153847 2017    <NA>    2010    PRIVATE 15    Don Va~ 1000 M~ 13
# ... with 11,641 more rows, 30 more variables: CONFIRMED_UNITS <chr>,
#   EVALUATION_COMPLETED_ON <chr>, SCORE <chr>, RESULTS_OF_SCORE <chr>,
#   NO_OF_AREAS_EVALUATED <chr>, ENTRANCE_LOBBY <chr>,
#   ENTRANCE_DOORS_WINDOWS <chr>, SECURITY <chr>, STAIRWELLS <chr>,
#   LAUNDRY_ROOMS <chr>, INTERNAL_GUARDS_HANDRAILS <chr>,
#   GARBAGE_CHUTE_ROOMS <chr>, GARBAGE_BIN_STORAGE_AREA <chr>, ELEVATORS <chr>,
```

```
#   STORAGE_AREAS_LOCKERS <chr>, INTERIOR_WALL_CEILING_FLOOR <chr>, ...
```

```
 # group_by(PROPERTY_TYPE)|>
  #mean(c(score<-as.integer(SCORE)))
  class(data$PROPERTY_TYPE)
```

```
[1] "character"
```

```
score<-as.numeric(data$SCORE)
mean(score)
```

```
[1] 73.77084
```

References