Thesis Outline

Max Callaghan

August 7, 2015

Outline

The general purpose of the paper is to use census data and energy consumption data to analyse household emissions. The data present multiple avenues of exploration for gaining more accurate insights into the drivers of household emissions and so for better formulation of policy interventions.

1. 3-Dimensional Fixed Effects Model

Household data is available at a high level of geographical granularity, and, since the release of Census data from 2011, in two timesteps (2001 and 2011). The paper will aim to produce a 3 way fixed effects model, to look at drivers of changes in emissions in geographical areas over time.

2. Fixed Effect Tree Regression Model

Building on the tree regression model in "A Spatial Typology of Human Settlements and their CO2 Emissions in England", the project will use the two points in time - 2001 and 2011 - to develop a fixed effect tree regression model which should shed light on how emissions drivers like income and population density affect changes in emissions differently in different area typologies.

3. Additional data

In addition to Census data, data, if available, will be collected on local authority environmental regulation/the powers of local authorities to regulate on issues that may affect household emissions. This analysis will help to examine the influence of specific policy interventions, where these can be identified, and the overall regulatory environment on household emissions.

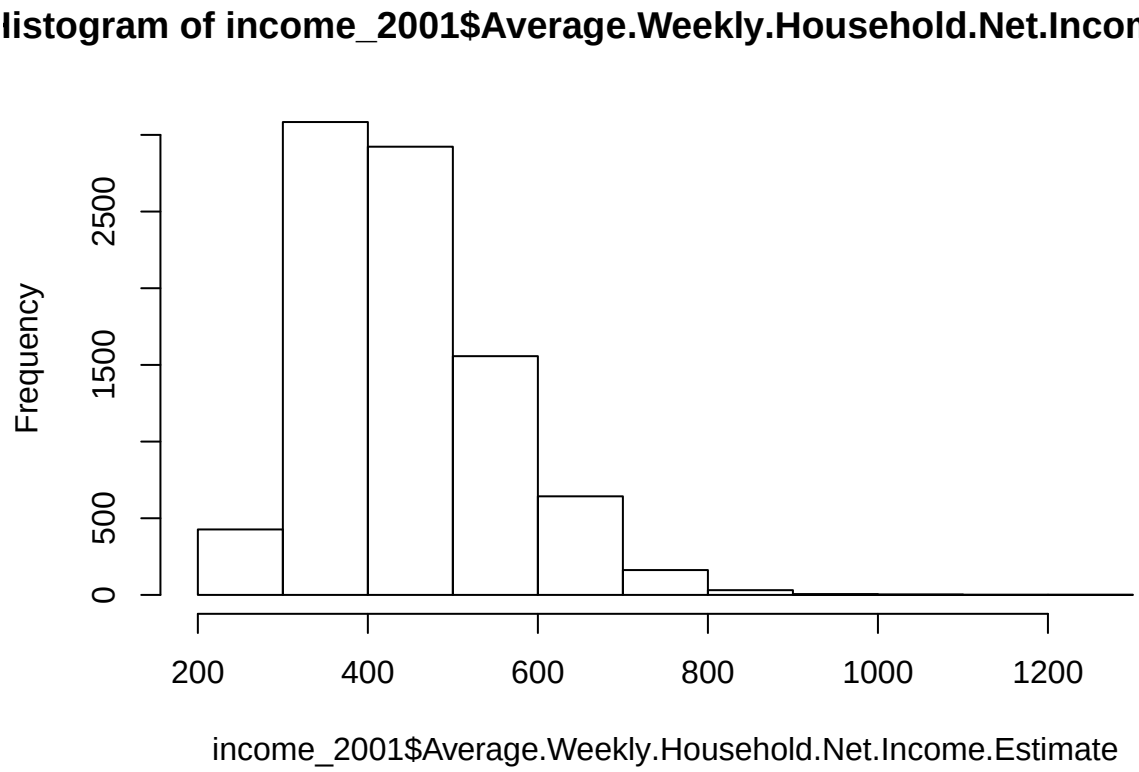
A more detailed profile of emissions may be gathered from smart meter data.

Variables and Indicators

Income

Data are available at ward level for Model-Based Income estimates.

```
library(tidyr)
library(dplyr)
library(stringr)

#-- data from http://www.neighbourhood.statistics.gov.uk/dissemination/instanceSelection.do?JSAllowed=t
income_2001 <- read.csv("income_2001/D120301_1153_2003Admin_WARD.CSV", header=TRUE,skip=5)
hist(income_2001$Average.Weekly.Household.Net.Income.Estimate)
```



We can convert these to MSOA level using this online tool

Data from 2011 was expected in the second quarter of 2015. Link

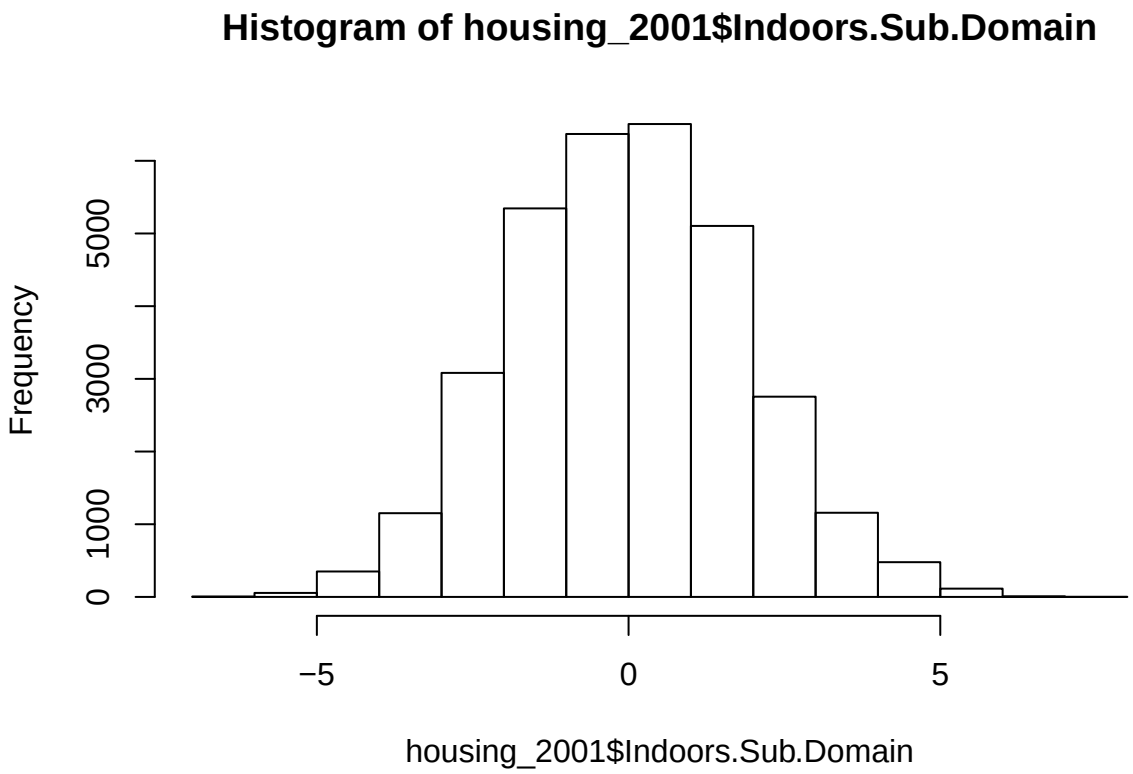
Heating Degree Days

Heating degree days are only available up to 2006, at UKCIP but can be calculated from UKCP09 monthly data using the formula here. Data will be available once registration has been approved.

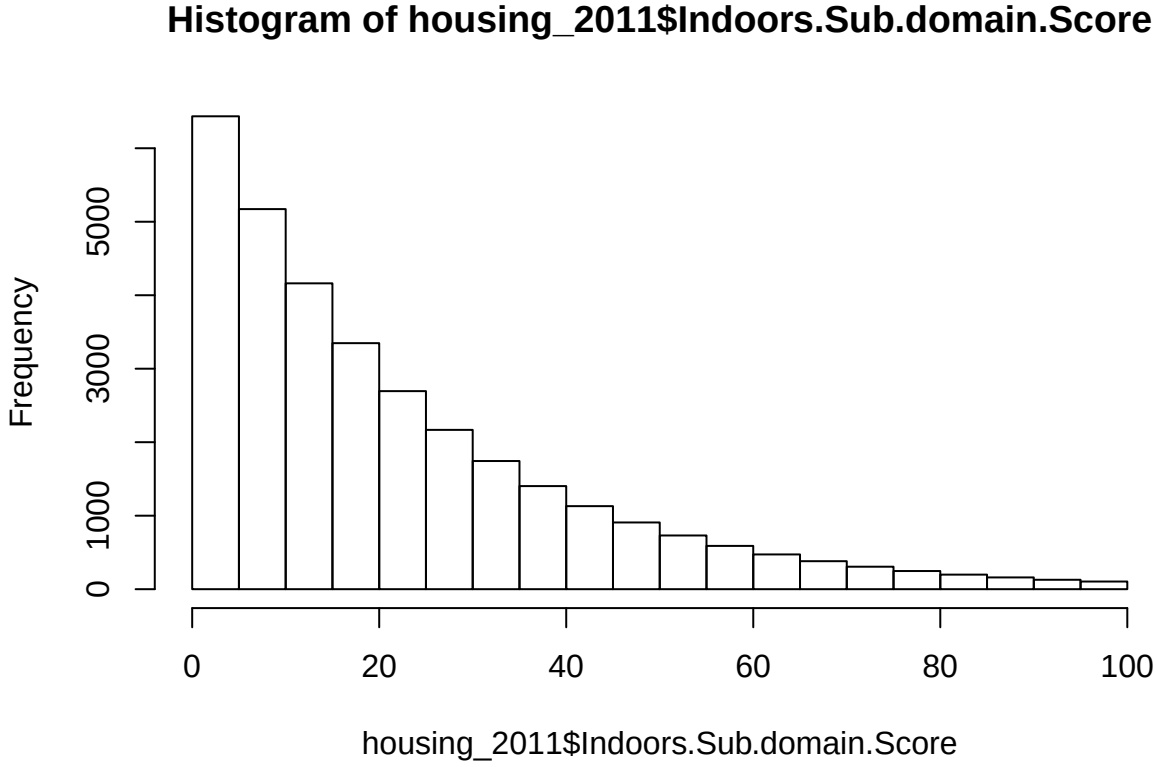
Central heating and housing in poor condition

These sub-indicators of the indoors sub-domain of the living environment domain of the index of multiple deprivation are not available at a disaggregated level. They could be recreated from their sources in housing surveys.

```
#-- data from http://webarchive.nationalarchives.gov.uk/20100410180038/http://www.communities.gov.uk/ar

housing_2001 <- read.csv("living environment/living_environment_2001.csv", header=TRUE)
hist(housing_2001$Indoors.Sub.Domain)
```




```
#-- https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6879/1871567.xls --#
housing_2011 <- read.csv("living environment/living_environment.csv", header=TRUE)
hist(housing_2011$Indoors.Sub.domain.Score)
```



Need to find out what's going on here - will need recoding somehow.

Population Density

Area statistics for MSOA come from (I used area not covered by water) here

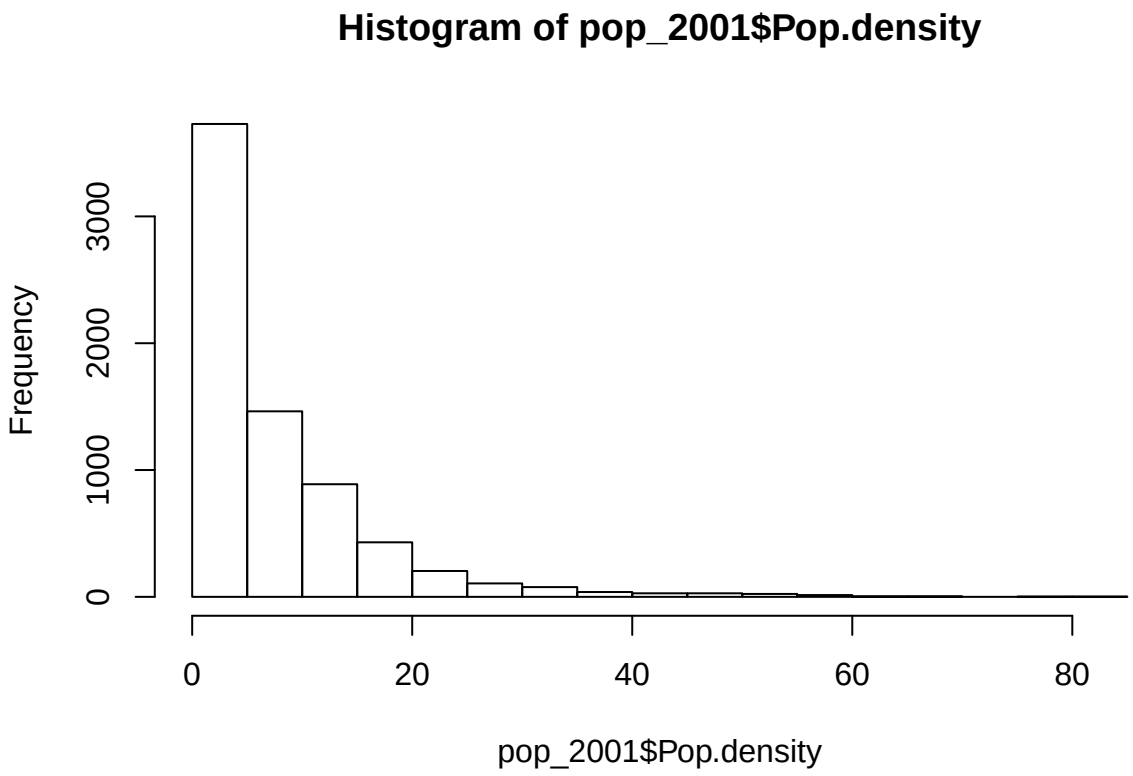
Population statistics come from here (2001) and here (2011)

```
area <- read.csv("msoas/area/MSOA11_LAD11_EW_SAM.csv", header=TRUE) %>%
  select(MSOA.Code=MSOA11CD,area=AREALHECT)

pop_2001 <- read.csv("msoas/mid_2001.csv", header=TRUE, skip=3) %>%
  select(MSOA.Code,MSOA.Name,Population = All.Ages) %>%
  mutate(Population = as.numeric(Population)) %>%
  left_join(area) %>%
  mutate(
    Pop.density = Population / area
  )
```

```
## Joining by: "MSOA.Code"
```

```
## Warning in left_join_impl(x, y, by$x, by$y): joining factors with different
## levels, coercing to character vector
```

```
hist(pop_2001$Pop.density,breaks=20)
```



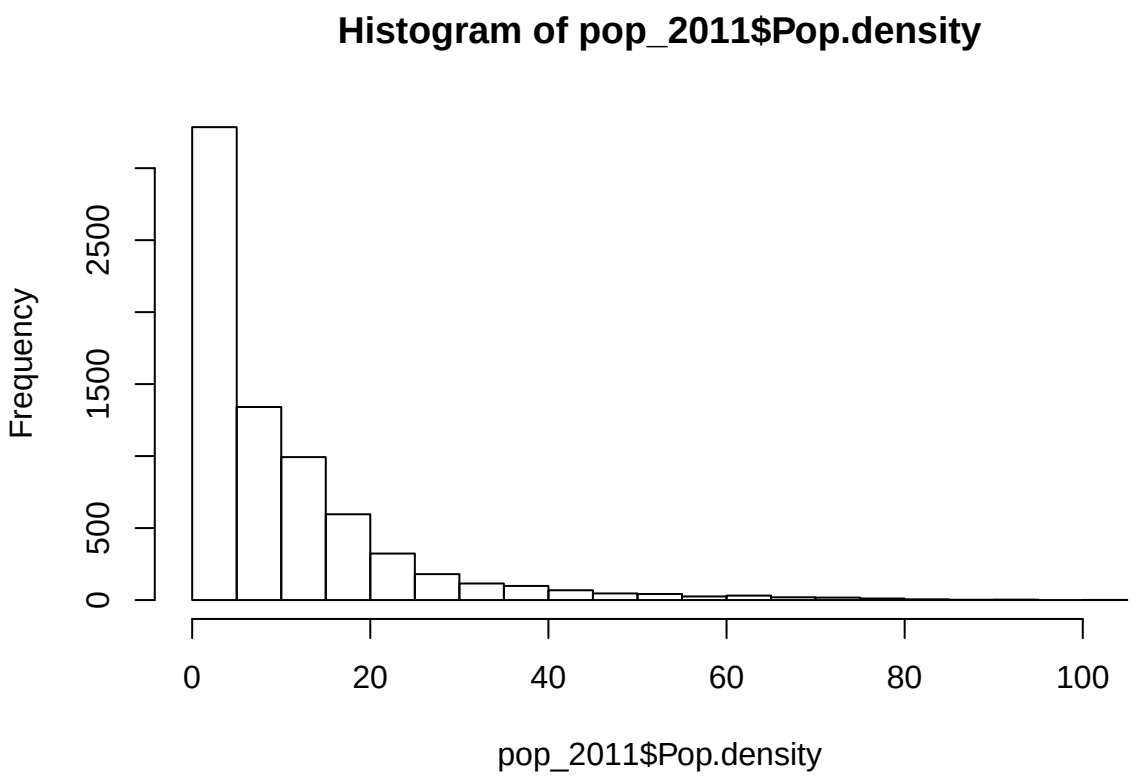

```
pop_2011 <- read.csv("msoas/mid_2011.csv", header=TRUE, skip=3)[,c(1,3,4)] %>%
  filter(X!="")

pop_2011 <- read.csv("msoas/mid_2011.csv", header=TRUE, skip=3) %>%
```

```
  select(MSOA.Code = Area.Codes,MSOA.Name=X,Population = All.Ages) %>%
  filter(MSOA.Name!="") %>%
  mutate(Population = as.numeric(Population)) %>%
  left_join(area) %>%
  mutate(
    Pop.density = Population / area
  )
```

```
## Joining by: "MSOA.Code"
```

```
## Warning in left_join_impl(x, y, by$x, by$y): joining factors with different
## levels, coercing to character vector
```

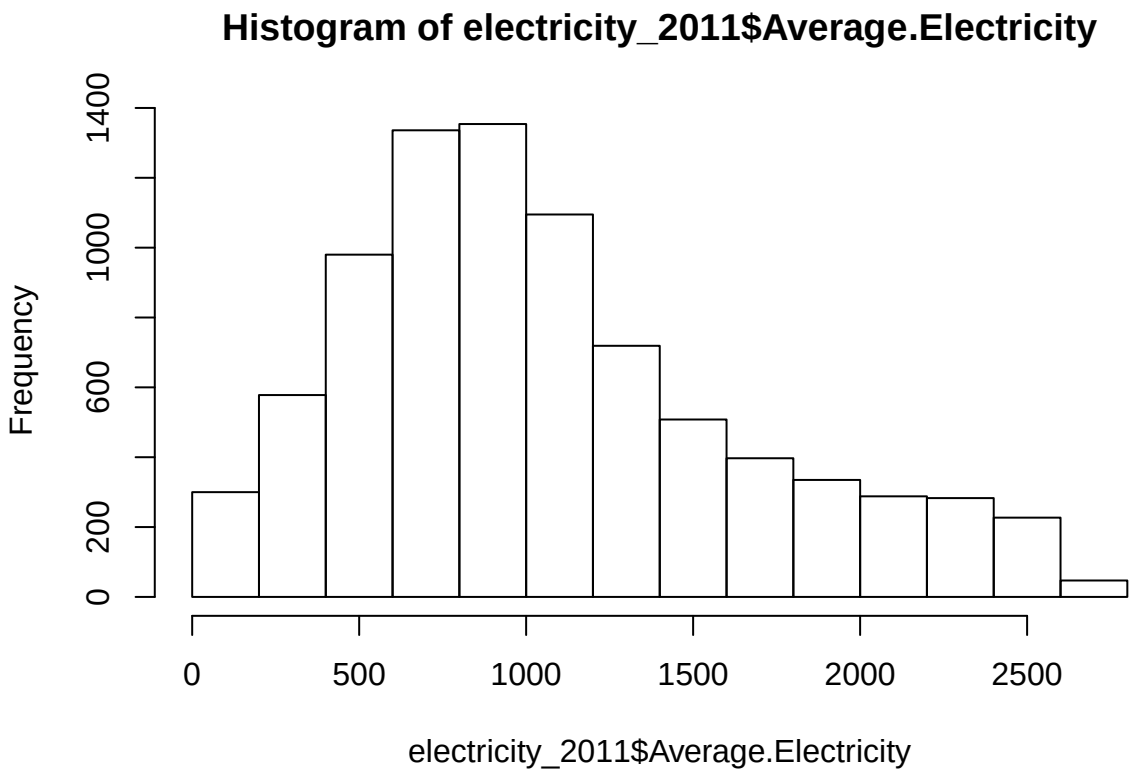
```
hist(pop_2011$Pop.density,breaks=20)
```



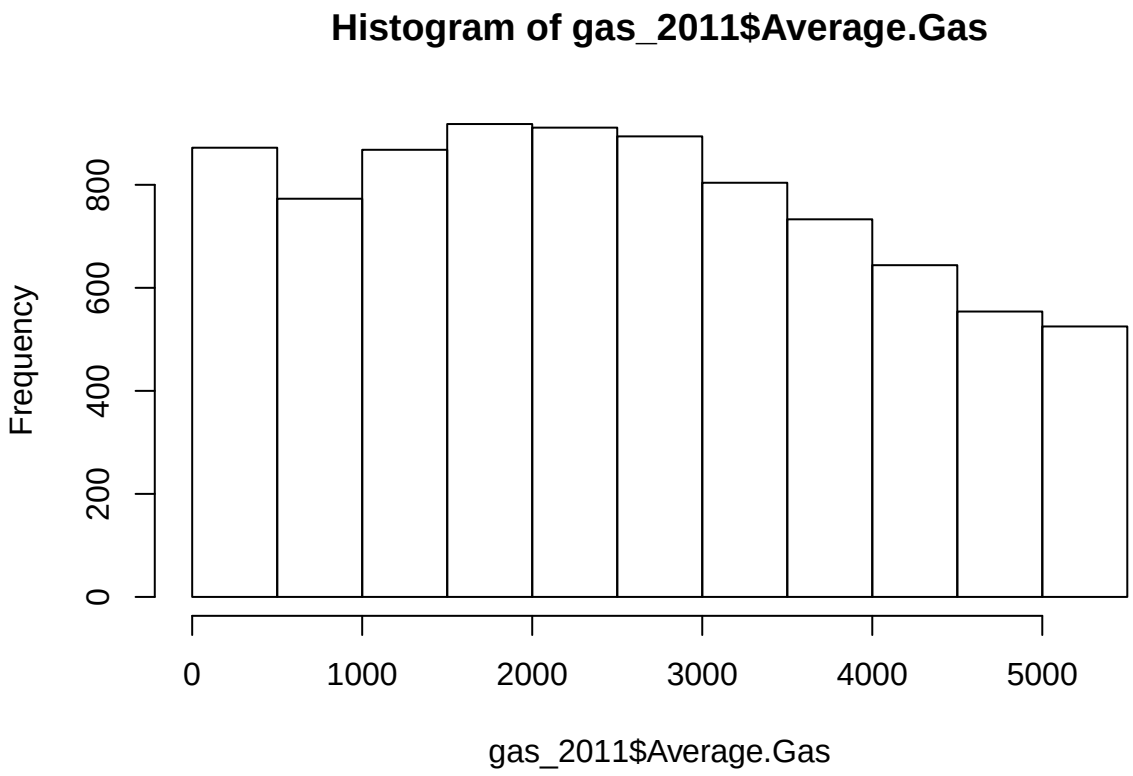
Energy consumption

Energy consumption statistics are available for 2011 - electricity and gas and for 2005. 2005 data needs a lot of rearranging.

```
electricity_2011 <- read.csv("electricity and gas/electricity/2011/MSOA_domestic_electricity_estimates_
  select(MSOA.Code = Middle.Layer.Super.Output.Area..MSOA..Code, Average.Electricity = Average.domestic
  mutate(
    Average.Electricity = as.numeric(Average.Electricity),
    MSOA.Code = str_trim(as.character(MSOA.Code))
        )
hist(electricity_2011$Average.Electricity)
```




```
gas_2011 <- read.csv("electricity and gas/gas/2011/MSOA_domestic_gas_estimates__2011.csv",skip=1) %>%
  select(MSOA.Code = Middle.Layer.Super.Output.Area..MSOA..Code, Average.Gas = Average.consumption..kWh
  mutate(
    Average.Gas = as.numeric(Average.Gas),
    MSOA.Code = str_trim(as.character(MSOA.Code))
        )
hist(gas_2011$Average.Gas)
```

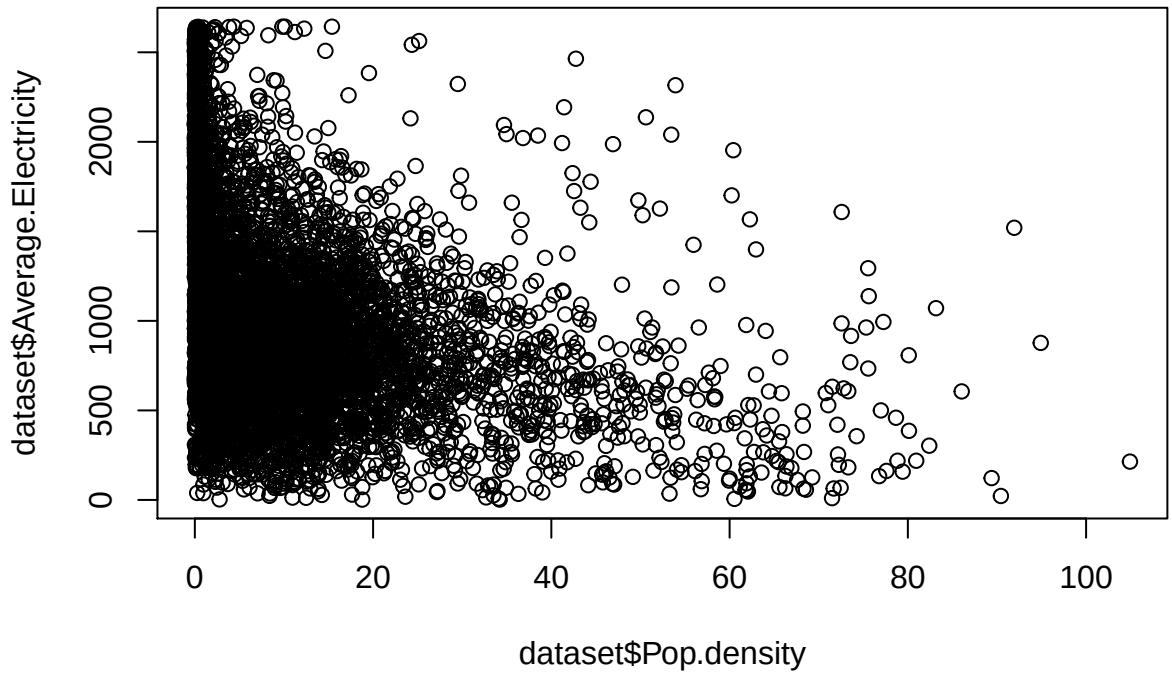


Sense test for data

```
dataset <- pop_2011 %>%
  left_join(electricity_2011)
```

```
## Joining by: "MSOA.Code"
```

```
plot(dataset$Pop.density,dataset$Average.Electricity)
```


dataset$Average.Electricity
0
500
1000
2000
0
20
40
60
80
100
dataset$Pop.density