initial explorations

Jon\_Minton

Friday, January 23, 2015

# Introduction

This document will show the results of some initial explorations of the relationship between pollution estimates at datazones and other variables available on the SNS website.

# Run

## Loading required package: repmis  
## Loading required package: plyr  
## Loading required package: dplyr  
##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:plyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize  
##   
## The following object is masked from 'package:stats':  
##   
## filter  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union  
##   
## Loading required package: tidyr  
## Loading required package: ggplot2

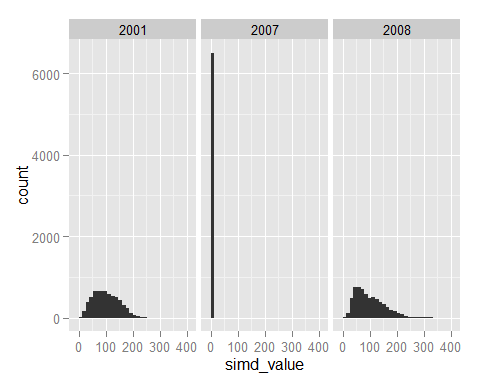
## SIMD correlations between years

# simd vars  
  
simd\_2009 <- source\_DropboxData(  
 file="simd\_2009.csv",  
 key="ghiu8n9db6rch9y"   
 ) %>% tbl\_df()

## Downloading data from: https://dl.dropboxusercontent.com/s/ghiu8n9db6rch9y/simd\_2009.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## 7e7197a527ecee20d8d80db0ec53c1b23cc18c2f

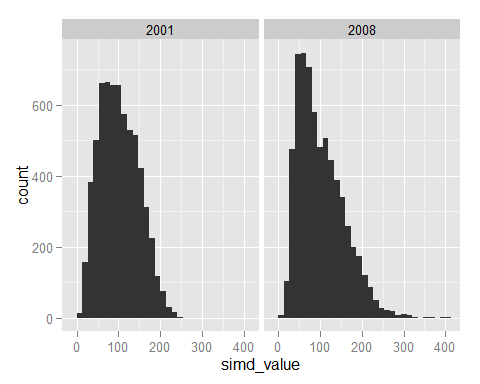
simd\_2009 <- simd\_2009 %>% gather("simd\_type", "simd\_value", c(2,4,5))  
  
qplot(data=simd\_2009, x=simd\_value, facets = . ~ year )

## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.



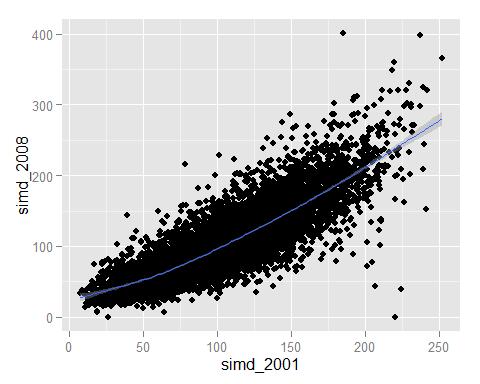
# clear no real values for 2007, so drop  
  
simd\_2009 <- simd\_2009 %>% filter(year !=2007)  
  
qplot(data=simd\_2009, x=simd\_value, facets = . ~ year)

## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.



# correlation between 2001 and 2008 simd?  
  
simd\_2009 <- simd\_2009 %>%   
 select(-simd\_type) %>%   
 group\_by(datazone) %>%   
 filter(!is.na(simd\_value)) %>%   
 spread(key=year, value=simd\_value)   
  
names(simd\_2009)[2:3] <- c("simd\_2001", "simd\_2008")  
  
qplot(data=simd\_2009,  
 x=simd\_2001, y=simd\_2008  
) + stat\_smooth()

## geom\_smooth: method="auto" and size of largest group is >=1000, so using gam with formula: y ~ s(x, bs = "cs"). Use 'method = x' to change the smoothing method.

 ## Pollution and SIMD

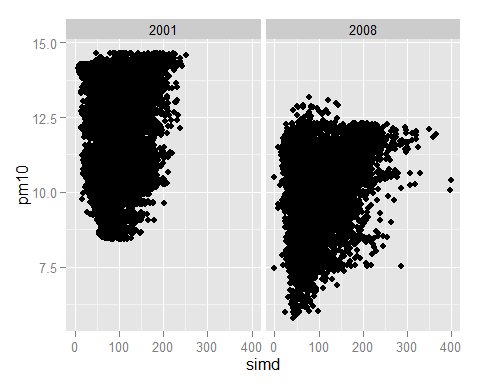
# OK - now pollution for same years  
  
pollution <- source\_DropboxData(  
 file="pollution\_by\_datazone.csv",  
 key="dziu2mknif22rmp"   
) %>% tbl\_df()

## Downloading data from: https://dl.dropboxusercontent.com/s/dziu2mknif22rmp/pollution\_by\_datazone.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## c70809779ad1a59de922d6a57f6a0613b0683c11

#is pm10 available for both years?  
# yes  
# pm2.5 first available from 2003  
  
# keep pm10, datazone, year  
pm10 <- pollution %>%   
 filter(year== 2001 | 2008) %>%  
 select(datazone, year, pm10)  
  
simd <- simd\_2009 %>% gather(key=year, value=simd, -datazone)  
# move simd back to long format for better merging  
simd$year <- as.character(simd$year)  
simd$year[simd$year=="simd\_2001"] <- "2001"  
simd$year[simd$year=="simd\_2008"] <- "2008"  
simd$year <- as.numeric(simd$year)  
  
joined <- simd %>% inner\_join(pm10)

## Joining by: c("datazone", "year")

qplot(  
 x=simd,  
 y=pm10,  
 facets = . ~ year,  
 data=joined  
 )



# main differences: overall rates from 2001-2008  
# range of simds increased (different categories?)  
# no obvious association in either year   
  
lm(pm10 ~ simd, data=joined) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ simd, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.0598 -1.0350 -0.0441 1.0735 3.8046   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.060e+01 3.382e-02 313.26 <2e-16 \*\*\*  
## simd 4.919e-03 2.972e-04 16.55 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.713 on 13008 degrees of freedom  
## Multiple R-squared: 0.02063, Adjusted R-squared: 0.02055   
## F-statistic: 274 on 1 and 13008 DF, p-value: < 2.2e-16

lm(pm10 ~ simd, data=joined, subset=year==2001) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ simd, data = joined, subset = year == 2001)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6484 -1.3440 0.0997 1.3748 2.9183   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.149e+01 4.775e-02 240.6 <2e-16 \*\*\*  
## simd 4.785e-03 4.274e-04 11.2 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.59 on 6503 degrees of freedom  
## Multiple R-squared: 0.01891, Adjusted R-squared: 0.01876   
## F-statistic: 125.3 on 1 and 6503 DF, p-value: < 2.2e-16

lm(pm10 ~ simd, data=joined, subset=year==2008) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ simd, data = joined, subset = year == 2008)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.1802 -0.8626 0.3404 1.0333 3.0555   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.6942075 0.0352692 274.86 <2e-16 \*\*\*  
## simd 0.0051332 0.0003044 16.86 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.341 on 6503 degrees of freedom  
## Multiple R-squared: 0.04189, Adjusted R-squared: 0.04174   
## F-statistic: 284.3 on 1 and 6503 DF, p-value: < 2.2e-16

## Income deprivation

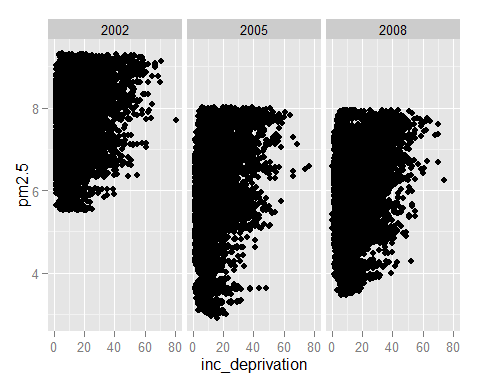
# Now let's look at income deprivation  
income\_deprivation <- source\_DropboxData(  
 file="income\_deprivation.csv",  
 key="tp66cm0xppfh07b"   
) %>% tbl\_df() %>%   
 filter(!is.na(CS.incdeprived)) %>%   
 rename(inc\_deprivation=CS.incdeprived)

## Downloading data from: https://dl.dropboxusercontent.com/s/tp66cm0xppfh07b/income\_deprivation.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## 519c70a285dc0bc86b035a3a90f686cde8a7f6ac

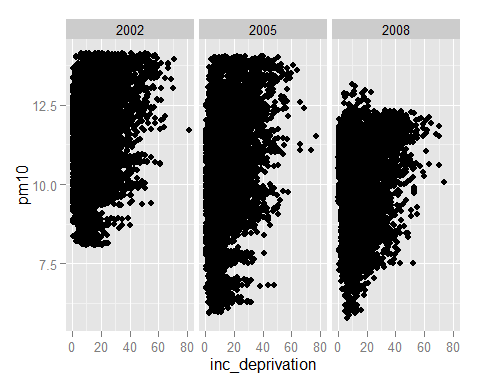
# from 2002 to 2008  
  
# do a left join with income\_deprivation as first argument  
  
joined <- income\_deprivation %>% left\_join(pollution)

## Joining by: c("datazone", "year")

# pm2.5 both years  
  
qplot(  
 x=inc\_deprivation,   
 y=pm2.5,  
 facets= . ~ year,  
 data=joined  
 )



# no obvious association   
  
qplot(  
 x=inc\_deprivation,   
 y=pm10,  
 facets= . ~ year,  
 data=joined  
)



# However, looking at it with linear models:  
  
lm(pm10 ~ inc\_deprivation, data=joined) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.8277 -1.0726 0.0898 1.1551 3.9237   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.018307 0.019157 523.0 <2e-16 \*\*\*  
## inc\_deprivation 0.034135 0.001019 33.5 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.632 on 19426 degrees of freedom  
## Multiple R-squared: 0.05462, Adjusted R-squared: 0.05457   
## F-statistic: 1122 on 1 and 19426 DF, p-value: < 2.2e-16

lm(pm10 ~ inc\_deprivation, data=joined, subset=year==2002) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation, data = joined, subset = year ==   
## 2002)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5662 -1.1556 0.0728 1.1859 3.0548   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.894664 0.028295 385.04 <2e-16 \*\*\*  
## inc\_deprivation 0.032244 0.001464 22.03 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.422 on 6433 degrees of freedom  
## Multiple R-squared: 0.07015, Adjusted R-squared: 0.07   
## F-statistic: 485.3 on 1 and 6433 DF, p-value: < 2.2e-16

lm(pm10 ~ inc\_deprivation, data=joined, subset=year==2005) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation, data = joined, subset = year ==   
## 2005)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.8729 -1.1561 -0.0835 1.3019 4.4643   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.270329 0.034934 265.37 <2e-16 \*\*\*  
## inc\_deprivation 0.050661 0.001949 25.99 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.726 on 6486 degrees of freedom  
## Multiple R-squared: 0.09433, Adjusted R-squared: 0.09419   
## F-statistic: 675.6 on 1 and 6486 DF, p-value: < 2.2e-16

lm(pm10 ~ inc\_deprivation, data=joined, subset=year==2008) %>% summary()

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation, data = joined, subset = year ==   
## 2008)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.2543 -0.8497 0.3489 1.0620 3.0339   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.941806 0.028410 349.94 <2e-16 \*\*\*  
## inc\_deprivation 0.017953 0.001486 12.08 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.355 on 6503 degrees of freedom  
## Multiple R-squared: 0.02196, Adjusted R-squared: 0.02181   
## F-statistic: 146 on 1 and 6503 DF, p-value: < 2.2e-16

# i.e. statistically significant association in each year  
# but adjusted R-squared values are VERY low

## SPRI (Pollutant sites)

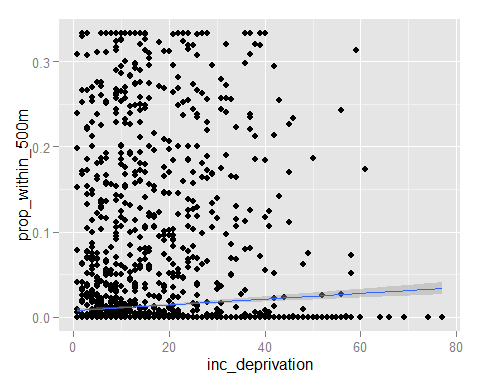
########## Now - spri sites   
# (special pollution releases inventory)  
  
  
  
# CS-CSdSPRIdpctpop2 Population within 500m of a site on the Scottish Pollutant Release Inventory  
# CS-CSdSPRIdpctpop4 Population within 1000m of a site on the Scottish Pollutant Release Inventory  
# CS-CSdSPRIdpctpop6 Population within 2000m of a site on the Scottish Pollutant Release Inventory  
# CS-CSdSPRIdpctpop8 Population over 2000m from a site on the Scottish Pollutant Release Inventory  
  
  
spri\_proximity <- source\_DropboxData(  
 file="proximity\_to\_a\_spri\_site.csv",  
 key="h9uqboghy1up0bj"   
) %>% tbl\_df() %>%  
 rename(  
 pop\_within\_500m=CS.CSdSPRIdpctpop2,  
 pop\_within\_1000m=CS.CSdSPRIdpctpop4,  
 pop\_within\_2000m=CS.CSdSPRIdpctpop6,  
 pop\_over\_2000m=CS.CSdSPRIdpctpop8  
 )

## Downloading data from: https://dl.dropboxusercontent.com/s/h9uqboghy1up0bj/proximity\_to\_a\_spri\_site.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## 8b080b1f948803d4caf39802ed8b37cff311b86f

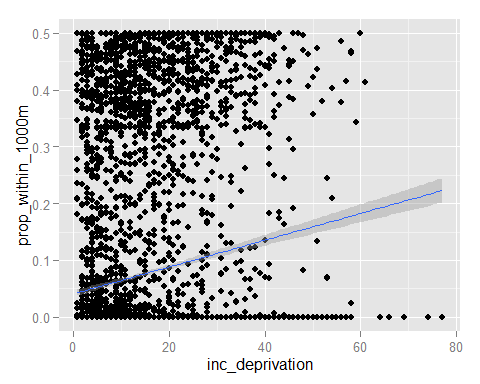
# both income deprivation and spri proximity are available in 2005  
  
joined <- income\_deprivation %>% inner\_join(spri\_proximity)

## Joining by: c("datazone", "year")

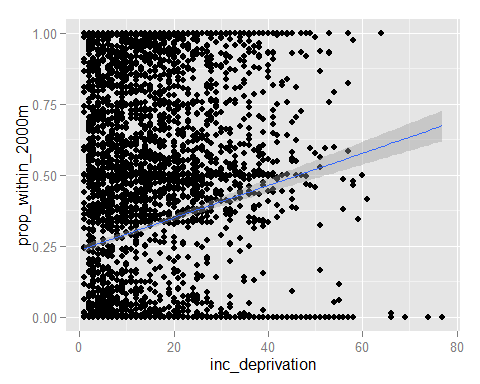
# really want proportions  
joined <- joined %>% mutate(  
 total\_pop = pop\_within\_500m + pop\_within\_1000m + pop\_within\_2000m + pop\_over\_2000m  
 )  
joined <- joined %>% transmute(  
 datazone=datazone, inc\_deprivation=inc\_deprivation,  
 prop\_within\_500m = pop\_within\_500m / total\_pop,   
 prop\_within\_1000m = pop\_within\_1000m / total\_pop,  
 prop\_within\_2000m = pop\_within\_2000m / total\_pop,  
 prop\_over\_2000m = pop\_over\_2000m / total\_pop  
 )  
  
qplot(  
 x=inc\_deprivation,  
 y= prop\_within\_500m,  
 data=joined  
 ) + stat\_smooth(method="lm")



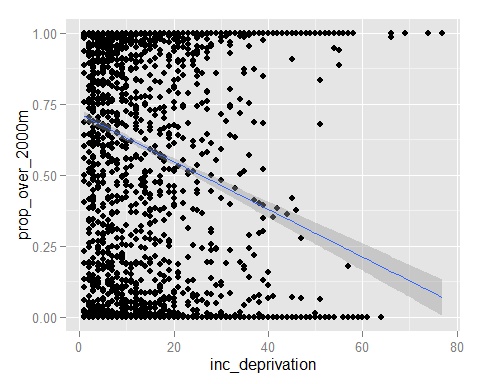
# very slight effect  
  
qplot(  
 x=inc\_deprivation,  
 y= prop\_within\_1000m,  
 data=joined  
) + stat\_smooth(method="lm")



# stat sig effect  
  
qplot(  
 x=inc\_deprivation,  
 y= prop\_within\_2000m,  
 data=joined  
) + stat\_smooth(method="lm")



qplot(  
 x=inc\_deprivation,  
 y= prop\_over\_2000m,  
 data=joined  
) + stat\_smooth(method="lm")



## Tenure type

# Now that we have some evidence of disproportionate siting, albeit only in 2005, we can look at   
# whether the tenure make-up of an area mediates this effect  
  
tenure\_households <- source\_DropboxData(  
 file="tenure\_households.csv",  
 key="kng5wc40le9kapj"   
) %>% tbl\_df() %>% select(  
 datazone, year,   
 all\_households=HO.allhouseholds,  
 council\_houses=HO.council,  
 rented\_from\_employer=HO.employ,  
 owned\_with\_mortgage=HO.ownmortloan,  
 owned\_outright=HO.ownoutright,  
 private\_rented=HO.privlet,  
 rented\_from\_relative=HO.relative,  
 shared\_ownership=HO.sharedown,  
 other\_social\_rented=HO.social  
 ) %>%   
 mutate(  
 social=council\_houses + other\_social\_rented,  
 rented=rented\_from\_employer + private\_rented+ rented\_from\_relative,  
 owned=owned\_with\_mortgage + owned\_outright + shared\_ownership  
 ) %>%  
 mutate(  
 council\_houses=council\_houses/all\_households,  
 rented\_from\_employer=rented\_from\_employer/all\_households,  
 owned\_with\_mortgage=owned\_with\_mortgage/all\_households,  
 owned\_outright=owned\_outright/all\_households,  
 private\_rented=private\_rented/all\_households,  
 rented\_from\_relative=rented\_from\_relative/all\_households,  
 shared\_ownership=shared\_ownership/all\_households,  
 other\_social\_rented=other\_social\_rented/all\_households,  
 social = social/all\_households,  
 rented = rented/all\_households,  
 owned=owned/all\_households  
 )

## Downloading data from: https://dl.dropboxusercontent.com/s/kng5wc40le9kapj/tenure\_households.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## 7cb8e90f51d1a3855e4ed4b9a7015a97d03282cd

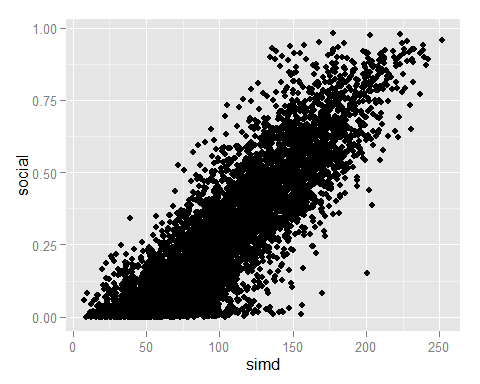
# then to pollution  
  
# tried income\_deprivation: no common year  
# but there is for simd. for 2001  
  
joined <- tenure\_households %>% inner\_join(simd)

## Joining by: c("datazone", "year")

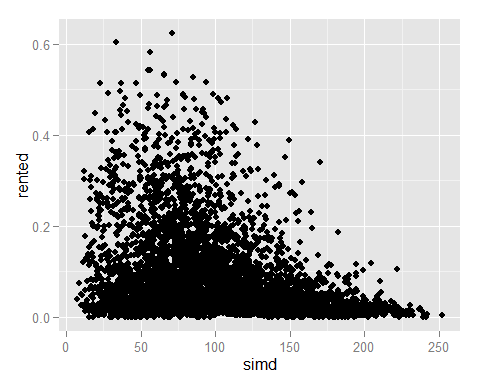
joined <- pollution %>% select(datazone, year, pm10) %>% inner\_join(joined)

## Joining by: c("datazone", "year")

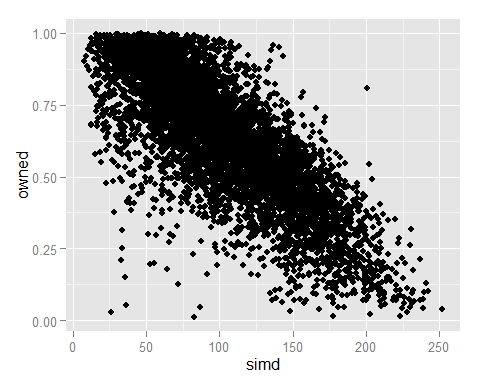
qplot(  
 x=simd,  
 y=social,  
 data=joined  
 )



# Strong positive association - but isn't this by design? (component of SIMD)  
# Not within the 2012 SIMD - not sure about older types of SIMD  
  
# What about rented?  
  
qplot(  
 x=simd,  
 y=rented,  
 data=joined  
)



# no clear assocaition  
  
qplot(  
 x=simd,  
 y=owned,  
 data=joined  
)



# strong negative association  
  
# This seems reassuring as to the choice of categories applied.

## External check on validity of datazone pollution estimates

# Found a variable on ambient air quality \_ There SHOULD be a strong correlation   
# between this and our pollution estimates!  
  
  
# CS-AQdNO2dpopwtd1 Air Quality 2002-2004 - Nitrogen Dioxide concentration - Population weighted NO2 concentration 2002-2004 ratio  
# CS-AQdPM10dpopwtd1 Air Quality 2002-2004 - PM10 concentration - Population weighted PM10 concentration 2002-2004 ratio  
  
ambient\_air\_quality <- source\_DropboxData(  
 file="ambient\_air\_quality.csv",  
 key="4gzgg8r3174xhfh"  
 ) %>% tbl\_df() %>%  
 select(  
 datazone, year,  
 pop\_weighted\_no2=CS.AQdNO2dpopwtd1,  
 pop\_weighted\_pm10=CS.AQdPM10dpopwtd1  
 )

## Downloading data from: https://dl.dropboxusercontent.com/s/4gzgg8r3174xhfh/ambient\_air\_quality.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## 3c7a9c47203e5d35da431458338a357533f88793

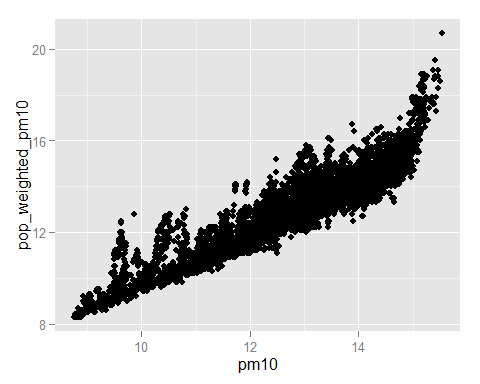
# If pop weighted we should also include population counts  
  
populations <- source\_DropboxData(  
 file="persons.csv",  
 key="vcz7qngb44vbynq"  
 ) %>% tbl\_df() %>% select(  
 datazone, year,  
 GR.hspeop,  
 GR.sapepeop  
 )

## Downloading data from: https://dl.dropboxusercontent.com/s/vcz7qngb44vbynq/persons.csv   
##   
## SHA-1 hash of the downloaded data file is:  
## 32a4513980642545d081f42cc28f7a5d8175f8be

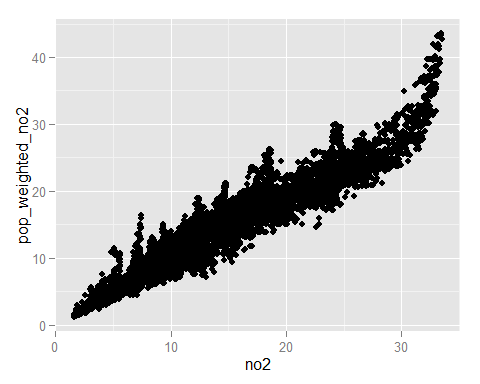
populations <- populations %>% mutate(  
 population\_count=ifelse(is.na(GR.sapepeop), GR.hspeop, GR.sapepeop)  
 ) %>% select(datazone, year, population\_count)  
  
joined <- populations %>%   
 inner\_join(ambient\_air\_quality) %>%   
 inner\_join(pollution)

## Joining by: c("datazone", "year")  
## Joining by: c("datazone", "year")

qplot(  
 x=pm10,  
 y=pop\_weighted\_pm10,  
 data=joined  
 )



# R2 of 0.84  
# Suggests we did something similar to what was done before  
  
qplot(  
 x=no2,  
 y=pop\_weighted\_no2,  
 data=joined  
)



# R2 of 0.91

## Explorations of tenure type and neighbourhood mix

##############################################################################################################  
# To do now:  
  
# regression: pm10 by income deprivation,   
# effect of adding prop social housing as a covariate  
  
joined <- pollution %>%   
 inner\_join(income\_deprivation)

## Joining by: c("year", "datazone")

# income deprivation is available for 2002, 2005, 2008  
# tenure\_households is available for 2001  
# pollution includes 2002 among other years  
  
p\_tmp <- pollution %>% filter(year==2002) %>% select(-year)  
i\_tmp <- income\_deprivation %>% filter(year==2002) %>% select(-year)  
t\_tmp <- tenure\_households %>% filter(year==2001) %>% select(-year)  
  
joined <- p\_tmp %>% inner\_join(i\_tmp) %>% inner\_join(t\_tmp)

## Joining by: "datazone"  
## Joining by: "datazone"

mod\_01 <- lm(  
 pm10 ~ inc\_deprivation,  
 data=joined)  
summary(mod\_01)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5662 -1.1556 0.0728 1.1859 3.0548   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.894664 0.028295 385.04 <2e-16 \*\*\*  
## inc\_deprivation 0.032244 0.001464 22.03 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.422 on 6433 degrees of freedom  
## Multiple R-squared: 0.07015, Adjusted R-squared: 0.07   
## F-statistic: 485.3 on 1 and 6433 DF, p-value: < 2.2e-16

mod\_02 <- mod\_01 %>% update(  
 . ~ . + social  
 )  
summary(mod\_02)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation + social, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.0992 -1.1400 0.1104 1.1796 3.2875   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.93291 0.02841 384.78 <2e-16 \*\*\*  
## inc\_deprivation 0.05673 0.00302 18.78 <2e-16 \*\*\*  
## social -1.43677 0.15532 -9.25 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.413 on 6432 degrees of freedom  
## Multiple R-squared: 0.08236, Adjusted R-squared: 0.08207   
## F-statistic: 288.6 on 2 and 6432 DF, p-value: < 2.2e-16

mod\_03 <- lm(pm10 ~ inc\_deprivation \* social,   
 data=joined)  
summary(mod\_03)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* social, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8550 -1.1518 0.1093 1.1768 3.3152   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.067301 0.040048 276.351 < 2e-16 \*\*\*  
## inc\_deprivation 0.040309 0.004585 8.791 < 2e-16 \*\*\*  
## social -1.706805 0.165143 -10.335 < 2e-16 \*\*\*  
## inc\_deprivation:social 0.027998 0.005890 4.754 2.04e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.411 on 6431 degrees of freedom  
## Multiple R-squared: 0.08557, Adjusted R-squared: 0.08514   
## F-statistic: 200.6 on 3 and 6431 DF, p-value: < 2.2e-16

mod\_social <- lm(  
 pm10 ~ inc\_deprivation \* social,   
 data=joined  
 )  
summary(mod\_social)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* social, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8550 -1.1518 0.1093 1.1768 3.3152   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.067301 0.040048 276.351 < 2e-16 \*\*\*  
## inc\_deprivation 0.040309 0.004585 8.791 < 2e-16 \*\*\*  
## social -1.706805 0.165143 -10.335 < 2e-16 \*\*\*  
## inc\_deprivation:social 0.027998 0.005890 4.754 2.04e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.411 on 6431 degrees of freedom  
## Multiple R-squared: 0.08557, Adjusted R-squared: 0.08514   
## F-statistic: 200.6 on 3 and 6431 DF, p-value: < 2.2e-16

mod\_rental <- lm(  
 pm10 ~ inc\_deprivation \* rented,  
 data=joined  
 )  
summary(mod\_rental)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* rented, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6023 -1.1210 0.0311 1.1471 3.7820   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.111029 0.036586 303.697 < 2e-16 \*\*\*  
## inc\_deprivation 0.025382 0.001799 14.110 < 2e-16 \*\*\*  
## rented -2.816219 0.331744 -8.489 < 2e-16 \*\*\*  
## inc\_deprivation:rented 0.104981 0.024116 4.353 1.36e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.413 on 6431 degrees of freedom  
## Multiple R-squared: 0.08245, Adjusted R-squared: 0.08202   
## F-statistic: 192.6 on 3 and 6431 DF, p-value: < 2.2e-16

mod\_owner <- lm(  
 pm10 ~ inc\_deprivation \* owned,  
 data=joined  
 )  
summary(mod\_owner)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* owned, data = joined)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.1229 -1.1082 0.0476 1.1295 3.8123   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.940085 0.128266 69.700 < 2e-16 \*\*\*  
## inc\_deprivation 0.074214 0.003337 22.241 < 2e-16 \*\*\*  
## owned 2.355937 0.148647 15.849 < 2e-16 \*\*\*  
## inc\_deprivation:owned -0.024812 0.006458 -3.842 0.000123 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.395 on 6431 degrees of freedom  
## Multiple R-squared: 0.1051, Adjusted R-squared: 0.1047   
## F-statistic: 251.8 on 3 and 6431 DF, p-value: < 2.2e-16

# now to normalise deprivation levels on a 0-1 scale  
  
j2 <- joined %>% mutate(inc\_deprivation=(inc\_deprivation - min(inc\_deprivation))/(max(inc\_deprivation)-min(inc\_deprivation)))  
  
# This \*might\* be helpful for understanding the relative importance of the interaction  
# between household type proportions and income deprivation  
  
norm\_social <- lm(  
 pm10 ~ inc\_deprivation \* social,   
 data=j2  
)  
summary(norm\_social)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* social, data = j2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.8550 -1.1518 0.1093 1.1768 3.3152   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.10761 0.03725 298.159 < 2e-16 \*\*\*  
## inc\_deprivation 3.22472 0.36683 8.791 < 2e-16 \*\*\*  
## social -1.67881 0.16321 -10.286 < 2e-16 \*\*\*  
## inc\_deprivation:social 2.23981 0.47118 4.754 2.04e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.411 on 6431 degrees of freedom  
## Multiple R-squared: 0.08557, Adjusted R-squared: 0.08514   
## F-statistic: 200.6 on 3 and 6431 DF, p-value: < 2.2e-16

norm\_rental <- lm(  
 pm10 ~ inc\_deprivation \* rented,  
 data=j2  
)  
summary(norm\_rental)

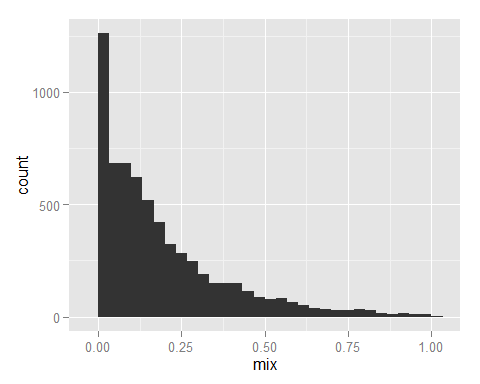
##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* rented, data = j2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6023 -1.1210 0.0311 1.1471 3.7820   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.13641 0.03522 316.230 < 2e-16 \*\*\*  
## inc\_deprivation 2.03055 0.14391 14.110 < 2e-16 \*\*\*  
## rented -2.71124 0.31314 -8.658 < 2e-16 \*\*\*  
## inc\_deprivation:rented 8.39849 1.92927 4.353 1.36e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.413 on 6431 degrees of freedom  
## Multiple R-squared: 0.08245, Adjusted R-squared: 0.08202   
## F-statistic: 192.6 on 3 and 6431 DF, p-value: < 2.2e-16

norm\_owner <- lm(  
 pm10 ~ inc\_deprivation \* owned,  
 data=j2  
)  
summary(norm\_owner)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* owned, data = j2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.1229 -1.1082 0.0476 1.1295 3.8123   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.0143 0.1257 71.688 < 2e-16 \*\*\*  
## inc\_deprivation 5.9371 0.2669 22.241 < 2e-16 \*\*\*  
## owned 2.3311 0.1471 15.852 < 2e-16 \*\*\*  
## inc\_deprivation:owned -1.9850 0.5167 -3.842 0.000123 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.395 on 6431 degrees of freedom  
## Multiple R-squared: 0.1051, Adjusted R-squared: 0.1047   
## F-statistic: 251.8 on 3 and 6431 DF, p-value: < 2.2e-16

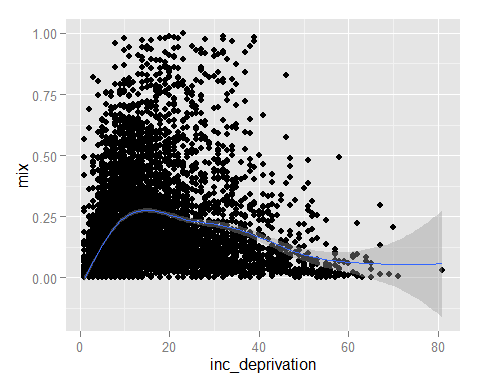
###################################################################################  
  
# As a crude measure of mix, let's use the produce of the three proportions   
# of household type, then normalised to a 0-1 scale  
  
j3 <- joined %>% mutate(mix=owned \*rented \* social) %>% mutate(mix =mix/max(mix))  
  
qplot(x=mix, data=j3)

## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.



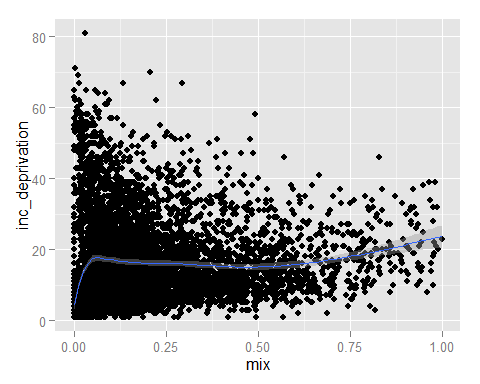
qplot(x=inc\_deprivation, y=mix, data=j3) + stat\_smooth()

## geom\_smooth: method="auto" and size of largest group is >=1000, so using gam with formula: y ~ s(x, bs = "cs"). Use 'method = x' to change the smoothing method.



qplot(y=inc\_deprivation, x=mix, data=j3) + stat\_smooth()

## geom\_smooth: method="auto" and size of largest group is >=1000, so using gam with formula: y ~ s(x, bs = "cs"). Use 'method = x' to change the smoothing method.



mix\_social <- lm(  
 pm10 ~ inc\_deprivation \* social \*mix,   
 data=j3  
)  
summary(mix\_social)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* social \* mix, data = j3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.182 -1.044 0.036 1.083 4.665   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.452476 0.046987 243.737 < 2e-16 \*\*\*  
## inc\_deprivation 0.034502 0.006540 5.276 1.37e-07 \*\*\*  
## social -1.525252 0.232814 -6.551 6.14e-11 \*\*\*  
## mix -5.165907 0.317404 -16.275 < 2e-16 \*\*\*  
## inc\_deprivation:social 0.021772 0.007493 2.906 0.00368 \*\*   
## inc\_deprivation:mix 0.228608 0.021548 10.609 < 2e-16 \*\*\*  
## social:mix 7.135437 1.246071 5.726 1.07e-08 \*\*\*  
## inc\_deprivation:social:mix -0.373202 0.047885 -7.794 7.54e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.367 on 6427 degrees of freedom  
## Multiple R-squared: 0.1421, Adjusted R-squared: 0.1412   
## F-statistic: 152.1 on 7 and 6427 DF, p-value: < 2.2e-16

mix\_rental <- lm(  
 pm10 ~ inc\_deprivation \* rented \* mix,  
 data=j3  
)  
summary(mix\_rental)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* rented \* mix, data = j3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.7419 -1.0472 0.0024 1.0750 4.1145   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.701071 0.044603 262.337 < 2e-16 \*\*\*  
## inc\_deprivation 0.012247 0.002198 5.571 2.64e-08 \*\*\*  
## rented -3.771139 0.509614 -7.400 1.54e-13 \*\*\*  
## mix -6.579323 0.309063 -21.288 < 2e-16 \*\*\*  
## inc\_deprivation:rented 0.151922 0.053809 2.823 0.00477 \*\*   
## inc\_deprivation:mix 0.194055 0.016562 11.717 < 2e-16 \*\*\*  
## rented:mix 20.418469 1.607278 12.704 < 2e-16 \*\*\*  
## inc\_deprivation:rented:mix -0.579144 0.089124 -6.498 8.74e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.352 on 6427 degrees of freedom  
## Multiple R-squared: 0.1605, Adjusted R-squared: 0.1596   
## F-statistic: 175.6 on 7 and 6427 DF, p-value: < 2.2e-16

mix\_owner <- lm(  
 pm10 ~ inc\_deprivation \* owned \* mix,  
 data=j3  
)  
summary(mix\_owner)

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* owned \* mix, data = j3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.2751 -1.0276 -0.0022 1.0493 3.9231   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.508838 0.188325 50.492 <2e-16 \*\*\*  
## inc\_deprivation 0.059834 0.004445 13.460 <2e-16 \*\*\*  
## owned 2.208733 0.201596 10.956 <2e-16 \*\*\*  
## mix 7.005914 0.649030 10.794 <2e-16 \*\*\*  
## inc\_deprivation:owned -0.017211 0.008294 -2.075 0.038 \*   
## inc\_deprivation:mix -0.233137 0.026640 -8.751 <2e-16 \*\*\*  
## owned:mix -16.639690 1.038579 -16.022 <2e-16 \*\*\*  
## inc\_deprivation:owned:mix 0.560095 0.057149 9.801 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.348 on 6427 degrees of freedom  
## Multiple R-squared: 0.1658, Adjusted R-squared: 0.1649   
## F-statistic: 182.5 on 7 and 6427 DF, p-value: < 2.2e-16

# Renove 3 way interaction to make interpretation simpler  
summary(lm(pm10 ~ inc\_deprivation\*social + mix, data=j3))

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* social + mix, data = j3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.522 -1.077 0.038 1.125 3.970   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.094404 0.039385 281.691 < 2e-16 \*\*\*  
## inc\_deprivation 0.073076 0.004990 14.643 < 2e-16 \*\*\*  
## social -1.316336 0.164249 -8.014 1.31e-15 \*\*\*  
## mix -1.578620 0.103447 -15.260 < 2e-16 \*\*\*  
## inc\_deprivation:social -0.020644 0.006606 -3.125 0.00179 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.386 on 6430 degrees of freedom  
## Multiple R-squared: 0.1175, Adjusted R-squared: 0.117   
## F-statistic: 214.1 on 4 and 6430 DF, p-value: < 2.2e-16

summary(lm(pm10 ~ inc\_deprivation\*owned + mix, data=j3))

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* owned + mix, data = j3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.219 -1.077 0.023 1.111 4.077   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.686867 0.152324 63.594 <2e-16 \*\*\*  
## inc\_deprivation 0.059031 0.003724 15.849 <2e-16 \*\*\*  
## owned 1.577482 0.171405 9.203 <2e-16 \*\*\*  
## mix -0.961509 0.107343 -8.957 <2e-16 \*\*\*  
## inc\_deprivation:owned -0.002612 0.006881 -0.380 0.704   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.387 on 6430 degrees of freedom  
## Multiple R-squared: 0.1162, Adjusted R-squared: 0.1156   
## F-statistic: 211.3 on 4 and 6430 DF, p-value: < 2.2e-16

summary(lm(pm10 ~ inc\_deprivation\*rented + mix, data=j3))

##   
## Call:  
## lm(formula = pm10 ~ inc\_deprivation \* rented + mix, data = j3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.4340 -1.0653 0.0286 1.0970 4.3876   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.351065 0.037896 299.530 < 2e-16 \*\*\*  
## inc\_deprivation 0.022952 0.001757 13.062 < 2e-16 \*\*\*  
## rented -2.459803 0.323726 -7.598 3.43e-14 \*\*\*  
## mix -2.542474 0.136455 -18.632 < 2e-16 \*\*\*  
## inc\_deprivation:rented 0.392672 0.028112 13.968 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.376 on 6430 degrees of freedom  
## Multiple R-squared: 0.1294, Adjusted R-squared: 0.1289   
## F-statistic: 239 on 4 and 6430 DF, p-value: < 2.2e-16