

# Gentrification and Health Disparities: An Integrated Analysis of Inequality in Philadelphia

Jenea Adams

Annan Timon

Sunday, April 30th, 2023

## Contents

<b>Packages</b>	<b>1</b>
<b>Executive Summary</b>	<b>1</b>
Abstract & Goal of the Study . . . . .	1
Data . . . . .	2
Brief Summary of Findings . . . . .	3
<b>EDA</b>	<b>3</b>
<b>Analysis</b>	<b>5</b>
Linear Model . . . . .	5
Model Selection . . . . .	6
Forward Selection . . . . .	7
Stroke . . . . .	7
COPD . . . . .	7
Backward Selection . . . . .	7
Stroke . . . . .	7
COPD . . . . .	7
Evaluation of Forward and Backward selection . . . . .	7
LASSO - predicting all diseases . . . . .	10
<b>Appendix I: Data Cleaning</b>	<b>15</b>
<b>Appendix II: Extended EDA</b>	<b>16</b>
<b>Appendix III: Data Dictionary</b>	<b>18</b>

## Packages

## Executive Summary

### Abstract & Goal of the Study

Gentrification is a process of affluent residents and businesses displacing existing low-income residents and businesses. Beyond prospects for so-called “urban renewal”, gentrification has real, tangible effects on the landscape and trajectories of existing communities who often don’t benefit from the changes of a neighborhood and are disenfranchised from participating in the growth of their area. Gentrification also has documented health effects on communities, such as shortened life expectancy, higher cancer rates, higher infant mortality, and cardiovascular diseases. Income inequality can be used to estimate gentrification rates. It can be

quantified by a [Gini index](#) which is a value from 0 to 1 indicating inequality in the dispersion of income in a given unit. **This study begins to to investigate a statistical framework for capturing the relationship between income inequality and health effects in Philadelphia, especially the historic Black Bottom (Figure 1), from integrated datasets.**

## Data

In this project, we use the following three nearly cleaned data:

**final\_data.RDS:** Census-Tract-level socioeconomic information that combines the following datqsets:

- **Food access:** The Neighborhood Food Retail dataset includes GEOID level assessments of food access relevant to distance and types of high produce grocery stores, as well as if that area is a high poverty area.
- **Hospital locations:** locations of hospitals by type of care provided. We chose this to understand the distribution and access to healthcare in each census block.
- **Heat Vulnerability:** scores and indicators for heat vulnerability by census block and prevalence of heat-related illnesses. This may or may not be related to access to green space and tree canopy. This data gives us an idea of environmental variables which contribute to health outcomes and quality of life for Philadelphia residents.
- **Affordable Housing:** locations of affordable and accessible housing projects recorded by the city. This data will provide information on the distribution of affordable housing options mapped to census blocks.
- **Philadelphia population metrics:** Demographic information of Philadelphia census blocks by race and ethnicity
- **Socioeconomic data:** income inequality calculated as a census tract-level Gini index.
- **Health data:** measures of prevalence of health measures such as cancer prevalence, access to health insurance, blood pressure, heart disease, and more by census tract.
- **Transitscore/Walkscore/Bikescore:** measures how walkable, well served by public transportation, and bikeable a location is

## Philadelphia's Black Bottom

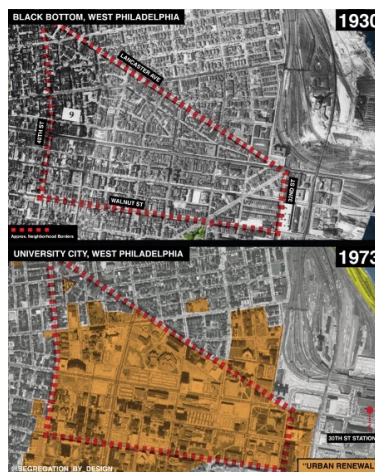
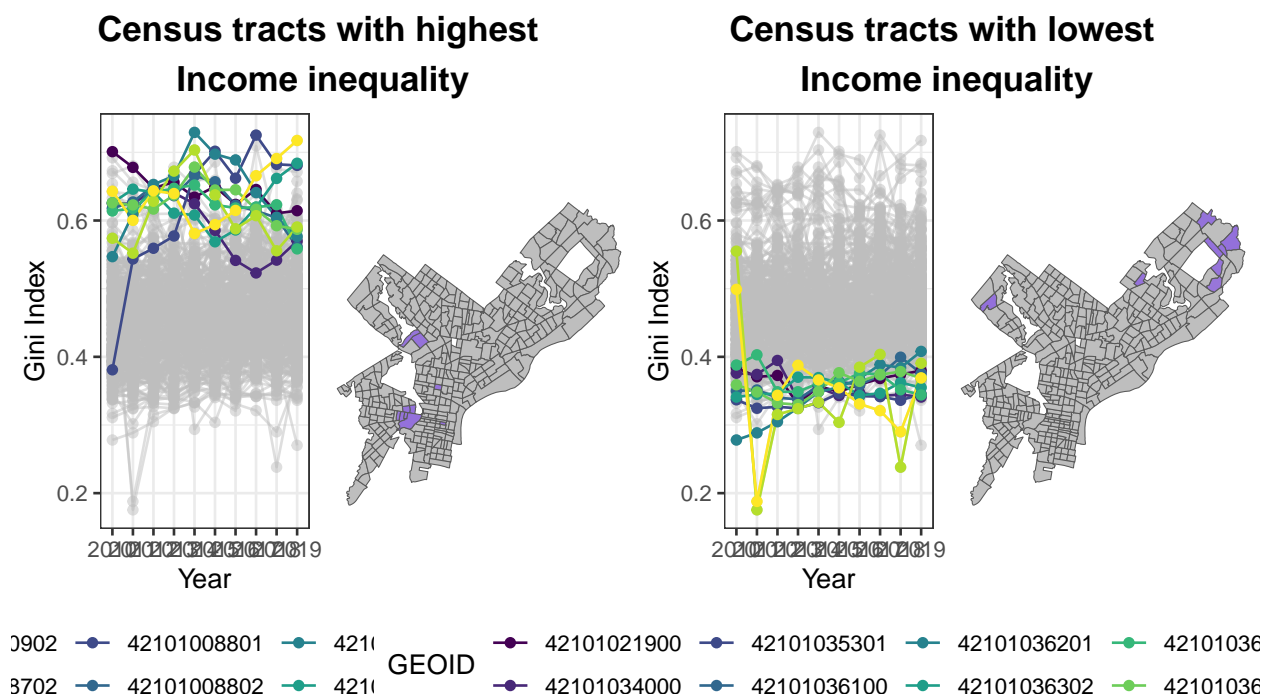


Figure 1: Philadelphia's Black Bottom from 1930 to 1973

**Gini:** measuring income inequality

Gini indexes have changed over time. Here we select the top and bottom 5 census tracts based on their mean Gini coefficient from 2011 to 2019. We see that the Black Bottom area has had a high mean gini index.

## |



## Brief Summary of Findings

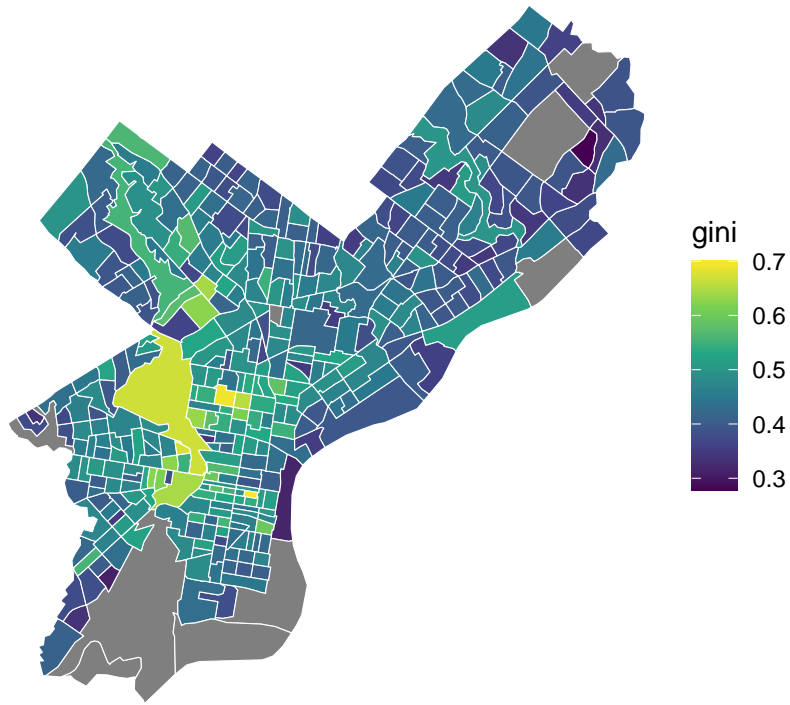
We found that there were strong associations between indicators of income inequality and adverse health outcomes in and closely bordering the footprint of Philadelphia's Black bottom. This association is particularly strong in in models used to predict COPD prevalence and Diabeates prevalence, diseases for which we see strong evidence of geospatial disparities in and around Philadelphia's Historic Black Bottom.

## EDA

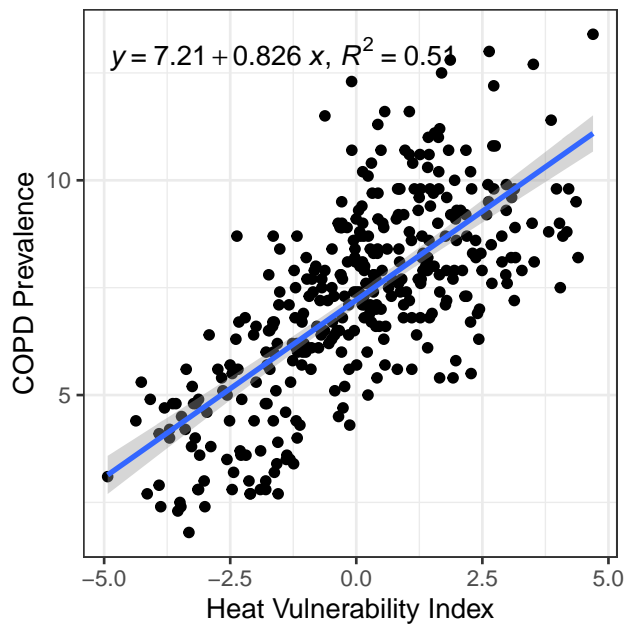
### looking specifically at 2010

Even though the last census concluded in 2020, we are using 2010 data because that was the year for the most robust and informative data available from our many sources.

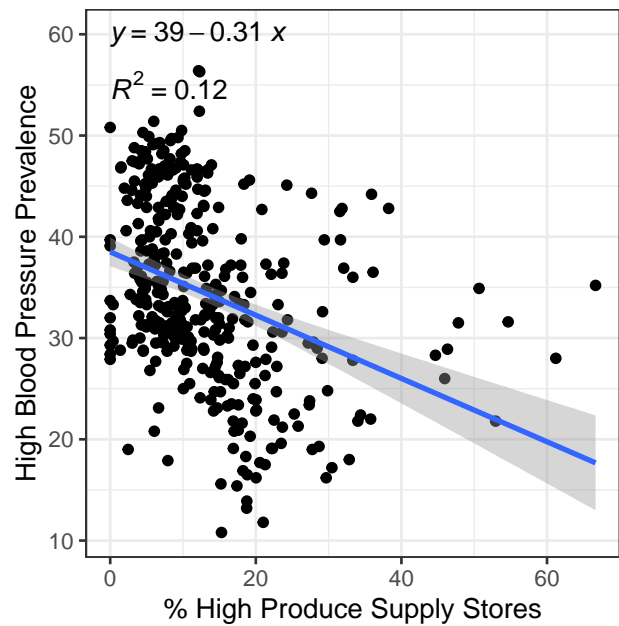
## |



Heat Vulnerability Index  
by COPD Prevalence

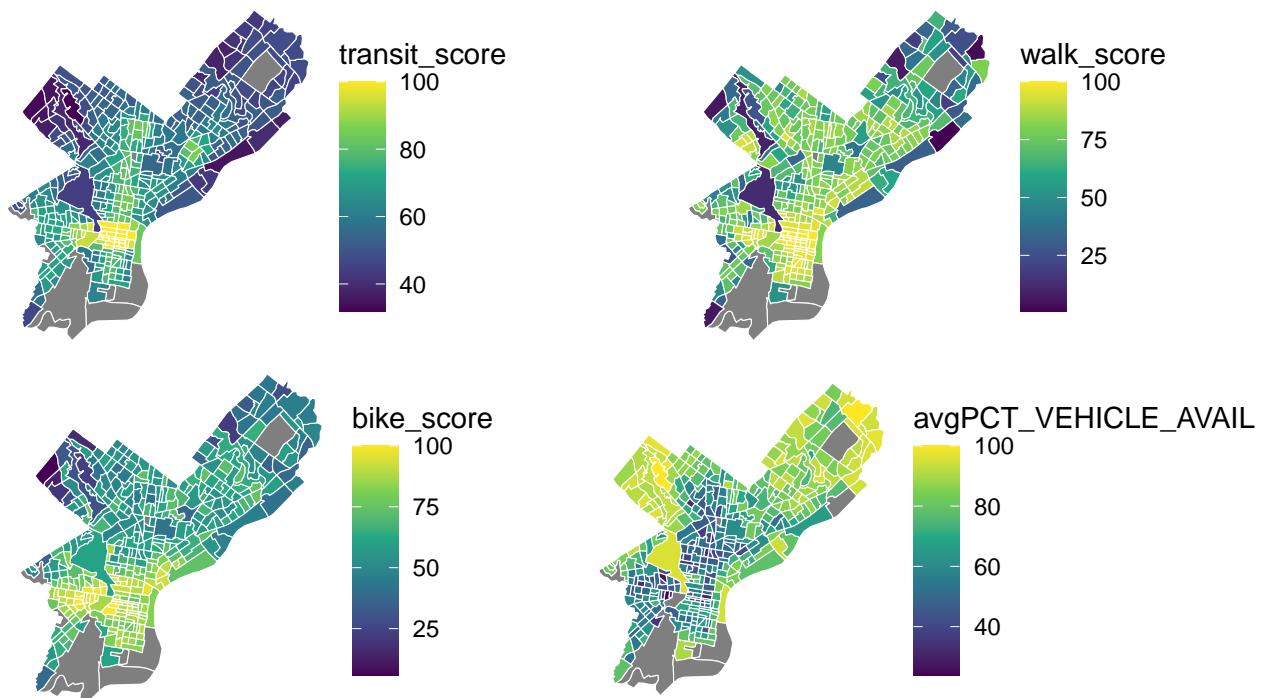


% of High Produce Supply Stores  
by High Blood Pressure Prevalence



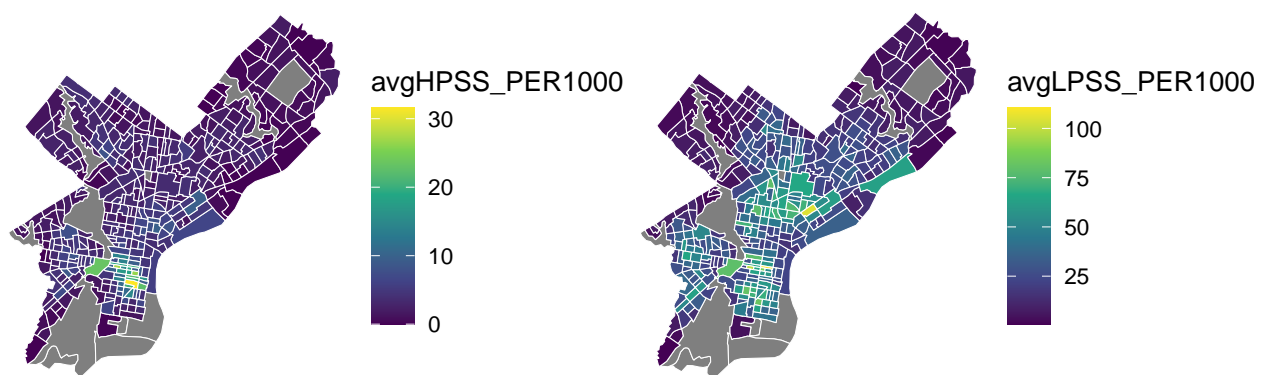
Transit and walks scores are lower in census tracts with higher (than average) inequality/Gini indexes.

## Utility of Public Transport, Walkability, Bikeability, and Car availability by Census Tracts



There's a wider distribution of low produce supply stores than high produce supply stores, which seem to be concentrated in the center of the city. We also have number of restaurants but don't include the analysis here for brevity

## High vs. Low Produce Supermarkets and Stores by Census Tracts



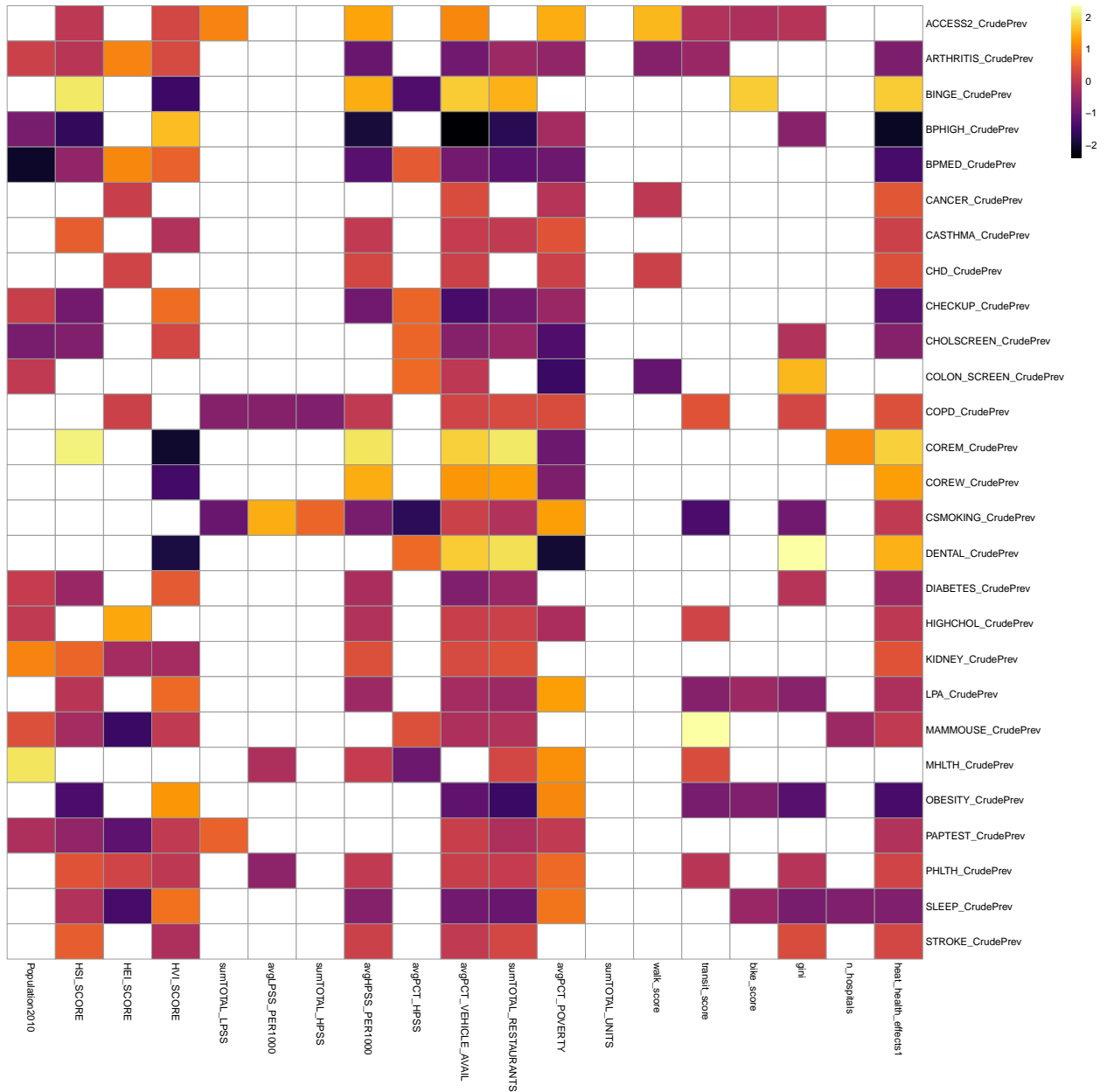
## Analysis

### Linear Model

Model data prep

```
## [1] 258 56
```

## [1] 111 56



Disease that with prevalences that increase with every increase in gini index: COLON\_SCREEN\_CrudePrev, COPD\_CrudePrev, DENTAL\_CrudePrev, STROKE\_CrudePrev Disease that with prevalences that decrease with every increase in gini index: ACCESS2\_CrudePrev, BPHIGH\_CrudePrev, CHOLSCREEN\_CrudePrev, CSMOKING\_CrudePrev, DIABETES\_CrudePrev, LPA\_CrudePrev, OBESITY\_CrudePrev, PHLTH\_CrudePrev, SLEEP\_CrudePrev

## Model Selection

Next we perform model selection on models to predict colon screening and COPD prevalence since they appear to have a strong positive correlation to gini index.

Prepare the data subsetting for model prediction of disease variables vs socioeconomic indicators

## Forward Selection

### Stroke

```
## [1] 1.058470e+03 1.867166e-28 1.614981e-28 1.276336e-28 1.117426e-28
## [6] 9.780608e-29 8.999488e-29 8.557649e-29 8.229538e-29 7.952600e-29
## [11] 7.809146e-29 7.615971e-29 7.518012e-29 7.402318e-29 7.303076e-29
## [16] 7.224276e-29 7.158858e-29 7.131089e-29 7.103575e-29 7.078254e-29
## [21] 7.074114e-29
```

### COPD

```
## [1] 21
```

## Backward Selection

### Stroke

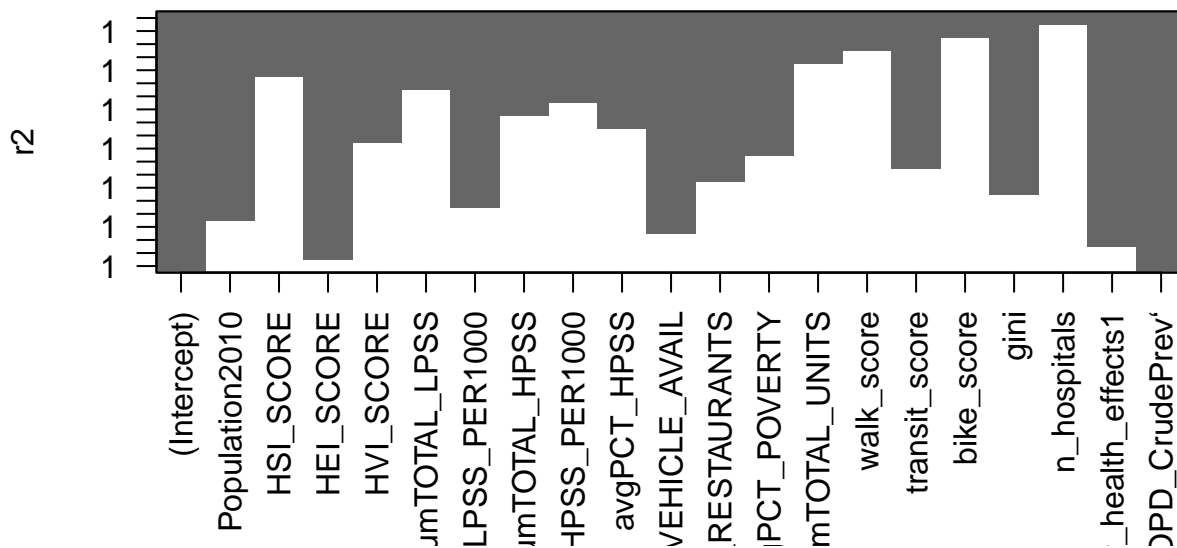
```
## [1] 21
```

### COPD

## Evaluation of Forward and Backward selection

We evaluate some of the outputs from `regsubsets` by looking for best performing subsets with high  $R^2$  values or low RSS (Residual Sum of Squares). Overall, it seems that forward and backward selection has the lowest RSS for the 20-variable model, but the plots show that there may be some variables that provide the highest  $R^2$ , such as heat exposure, vehicle availability, and experiencing health effects from heat exposure for the COPD prediction model.

```
plot(fit.backward.copd, scale = "r2")
```



### Stroke

```
which.min(summary(fit.forward.stroke)$rss)
```

```
## [1] 20
```

```
which.min(summary(fit.forward.stroke)$rss)
```

```
## [1] 20
```

```
coef(fit.forward.stroke, 20)
```

```
##              (Intercept)              Population2010
##      8.699494e-15          -6.122308e-20
##      HSI_SCORE              HEI_SCORE
##     -1.033855e-16          1.871907e-16
##      HVI_SCORE              sumTOTAL_LPSS
##      2.642465e-16          -2.469482e-18
##      avgLPSS_PER1000          sumTOTAL_HPSS
##      3.946810e-18          1.869696e-17
##      avgHPSS_PER1000          avgPCT_HPSS
##     -1.085258e-16          4.995873e-18
##      avgPCT_VEHICLE_AVAIL      sumTOTAL_RESTAURANTS
##     -5.194791e-17          -9.419641e-18
##      avgPCT_POVERTY          sumTOTAL_UNITS
##     -8.076808e-18          7.304433e-19
##      walk_score              transit_score
##     -5.399541e-18          -3.972348e-18
##      bike_score              gini
##     -9.688228e-19          -1.573594e-15
##      n_hospitals          heat_health_effects1
##     -1.012018e-16          -6.674627e-16
## `final_data$STROKE_CrudePrev`
##      1.000000e+00
```

```
which.min(summary(fit.backward.stroke)$rss)
```

```
## [1] 20
```

```
which.min(summary(fit.backward.stroke)$rss)
```

```
## [1] 20
```

```
coef(fit.backward.stroke, 20)
```

```
##              (Intercept)              Population2010
##      8.520469e-15          -9.142541e-20
##      HSI_SCORE              HEI_SCORE
##     -1.777926e-16          2.195129e-16
##      HVI_SCORE              sumTOTAL_LPSS
##      3.616050e-16          -1.893904e-18
##      avgLPSS_PER1000          sumTOTAL_HPSS
##      5.165871e-18          1.167587e-17
##      avgHPSS_PER1000          avgPCT_HPSS
##     -9.329250e-17          3.844135e-18
##      avgPCT_VEHICLE_AVAIL      sumTOTAL_RESTAURANTS
##     -4.530590e-17          -9.527703e-18
##      avgPCT_POVERTY          sumTOTAL_UNITS
##     -4.147210e-18          4.208901e-19
##      walk_score              transit_score
##     -5.079155e-18          -2.406180e-18
##      bike_score              gini
##     -6.324161e-19          -1.439776e-15
##      n_hospitals          heat_health_effects1
##     -7.699055e-17          -8.807395e-16
## `final_data$STROKE_CrudePrev`
```



```
## 1.000000e+00
```

The RSS is the smallest for the 20-variable model for Stroke in both forward and backward selection  
COPD

```
which.min(summary(fit.forward.copd)$rss)
```

```
## [1] 20
```

```
which.min(summary(fit.forward.copd)$rss)
```

```
## [1] 20
```

```
coef(fit.forward.copd, 20)
```

```
## (Intercept) Population2010
## 3.397340e-15 3.191191e-20
## HSI_SCORE HEI_SCORE
## -1.241787e-17 1.580660e-16
## HVI_SCORE sumTOTAL_LPSS
## 3.095724e-17 -4.975750e-19
## avgLPSS_PER1000 sumTOTAL_HPSS
## -5.322894e-19 8.596414e-18
## avgHPSS_PER1000 avgPCT_HPSS
## -3.001471e-17 -1.170297e-18
## avgPCT_VEHICLE_AVAIL sumTOTAL_RESTAURANTS
## -8.572170e-18 -1.127614e-18
## avgPCT_POVERTY sumTOTAL_UNITS
## 2.193157e-18 6.101153e-20
## walk_score transit_score
## -2.947037e-19 -2.388980e-18
## bike_score gini
## 3.005885e-19 -6.206644e-16
## n_hospitals heat_health_effects1
## -8.900563e-19 -2.411462e-16
## `final_data$COPD_CrudePrev`
## 1.000000e+00
```

```
which.min(summary(fit.backward.copd)$rss)
```

```
## [1] 20
```

```
which.min(summary(fit.backward.copd)$rss)
```

```
## [1] 20
```

```
coef(fit.backward.copd, 20)
```

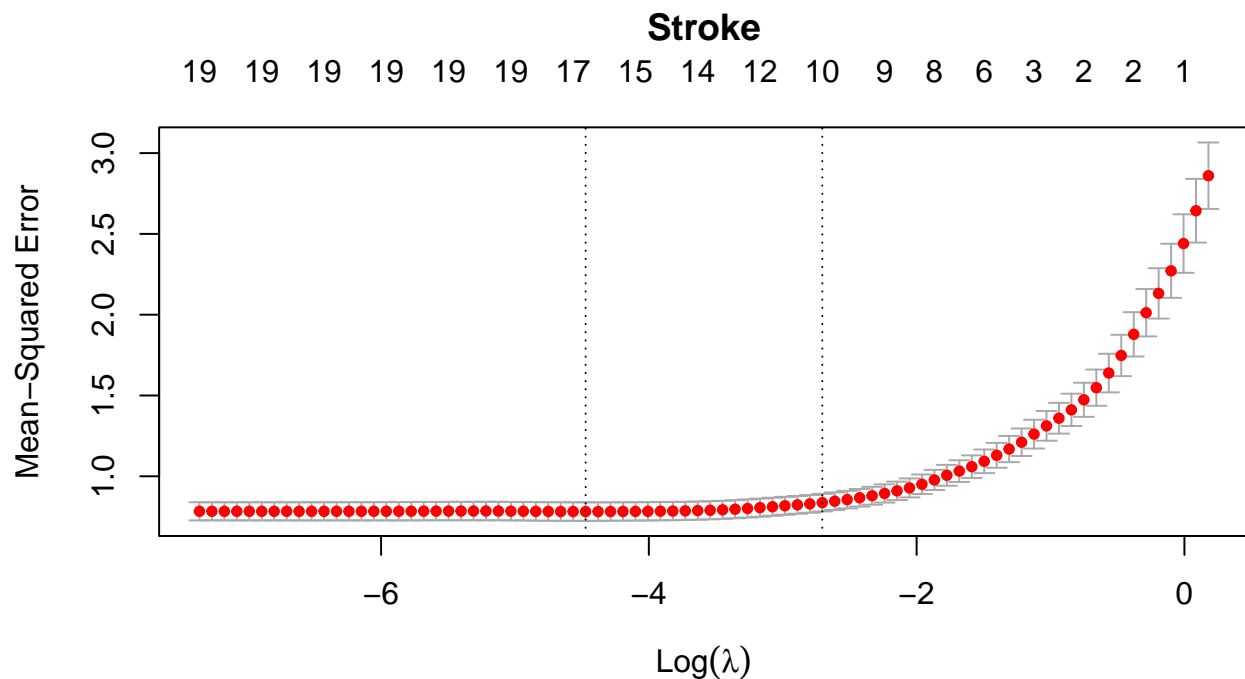
```
## (Intercept) Population2010
## 3.397340e-15 3.191191e-20
## HSI_SCORE HEI_SCORE
## -1.241787e-17 1.580660e-16
## HVI_SCORE sumTOTAL_LPSS
## 3.095724e-17 -4.975750e-19
## avgLPSS_PER1000 sumTOTAL_HPSS
## -5.322894e-19 8.596414e-18
## avgHPSS_PER1000 avgPCT_HPSS
## -3.001471e-17 -1.170297e-18
```

```
##      avgPCT_VEHICLE_AVAIL      sumTOTAL_RESTAURANTS
##      -8.572170e-18      -1.127614e-18
##      avgPCT_POVERTY      sumTOTAL_UNITS
##      2.193157e-18      6.101153e-20
##      walk_score      transit_score
##      -2.947037e-19      -2.388980e-18
##      bike_score      gini
##      3.005885e-19      -6.206644e-16
##      n_hospitals      heat_health_effects1
##      -8.900563e-19      -2.411462e-16
## `final_data$COPD_CrudePrev`
##      1.000000e+00
```

The RSS is the smallest for the 20-variable model for COPD in both forward and backward selection

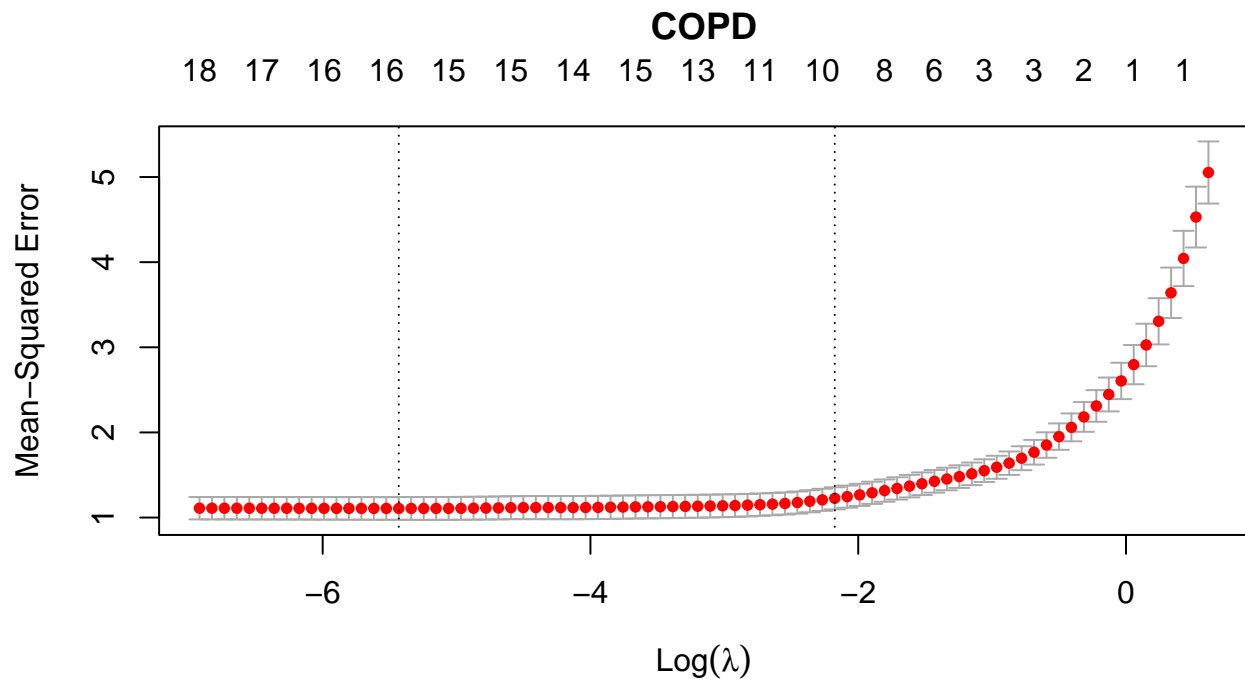
## LASSO - predicting all diseases

LASSO loop



```
##
## Call:
## lm(formula = f_new, data = final_data_sub)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7643 -0.5277 -0.0162  0.4666  3.0146
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.008e+01  6.235e-01  16.159  < 2e-16 ***
## Population2010 -1.345e-04  3.882e-05  -3.464  0.000596 ***
## HSI_SCORE     -2.329e-02  4.191e-02  -0.556  0.578728
## HEI_SCORE      6.603e-01  4.415e-02  14.955  < 2e-16 ***
```

```
## sumTOTAL_HPSS      1.229e-02  9.040e-03  1.359 0.174914
## avgHPSS_PER1000    -9.062e-02  3.183e-02  -2.847 0.004668 **
## avgPCT_VEHICLE_AVAIL -5.165e-02  3.692e-03 -13.990 < 2e-16 ***
## sumTOTAL_RESTAURANTS -1.036e-02  2.459e-03  -4.213 3.19e-05 ***
## walk_score         -3.959e-03  3.375e-03  -1.173 0.241500
## gini               -2.321e+00  8.384e-01  -2.768 0.005933 **
## heat_health_effects1 -9.634e-01  1.610e-01  -5.986 5.24e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8748 on 358 degrees of freedom
## Multiple R-squared:  0.7412, Adjusted R-squared:  0.734
## F-statistic: 102.5 on 10 and 358 DF,  p-value: < 2.2e-16
```



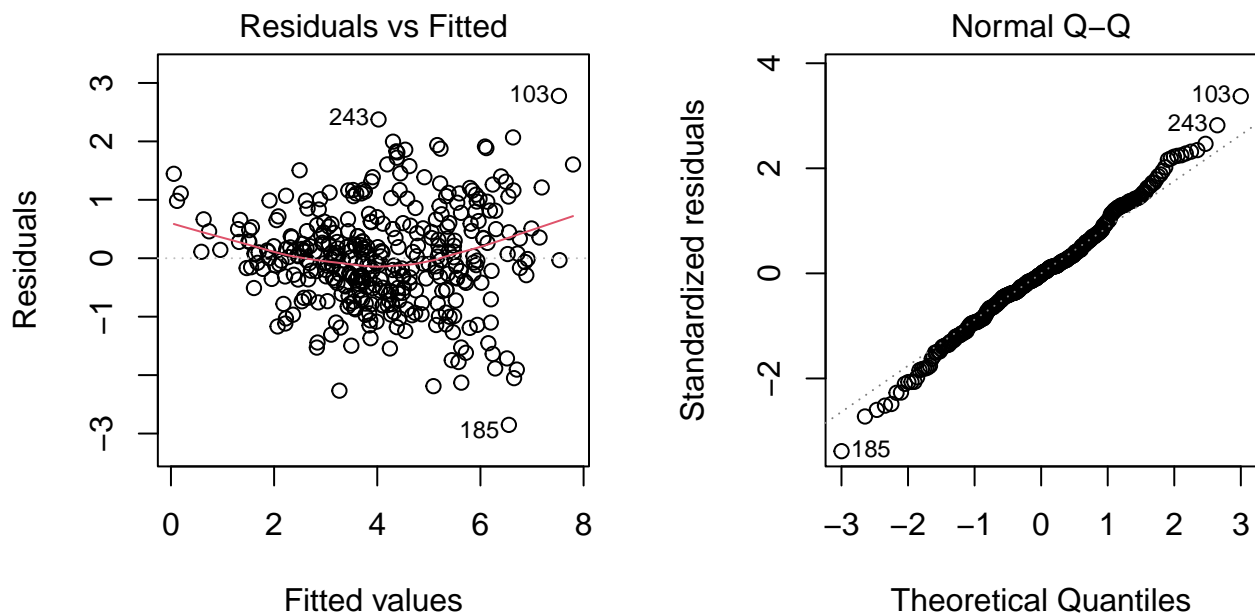
```
##
## Call:
## lm(formula = f_new, data = final_data_sub)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6771 -0.5763 -0.0015  0.6450  3.6077
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.449200   0.893747  15.048 < 2e-16 ***
## HEI_SCORE      0.878069   0.054300  16.171 < 2e-16 ***
## avgHPSS_PER1000 -0.036264   0.018637  -1.946 0.052457 .
## avgPCT_HPSS     -0.025461   0.006810  -3.739 0.000215 ***
## avgPCT_VEHICLE_AVAIL -0.042243   0.006187  -6.827 3.71e-11 ***
## sumTOTAL_RESTAURANTS -0.007578   0.002935  -2.582 0.010223 *
## avgPCT_POVERTY    0.012661   0.006041   2.096 0.036787 *
## walk_score     -0.001614   0.003903  -0.414 0.679468
```

```
## transit_score      -0.020714    0.006127   -3.381 0.000802 ***
## gini               -3.174678    0.990450   -3.205 0.001471 **
## heat_health_effects1 -0.663080    0.184491   -3.594 0.000371 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.03 on 358 degrees of freedom
## Multiple R-squared:  0.7971, Adjusted R-squared:  0.7914
## F-statistic: 140.6 on 10 and 358 DF,  p-value: < 2.2e-16
```

Stroke

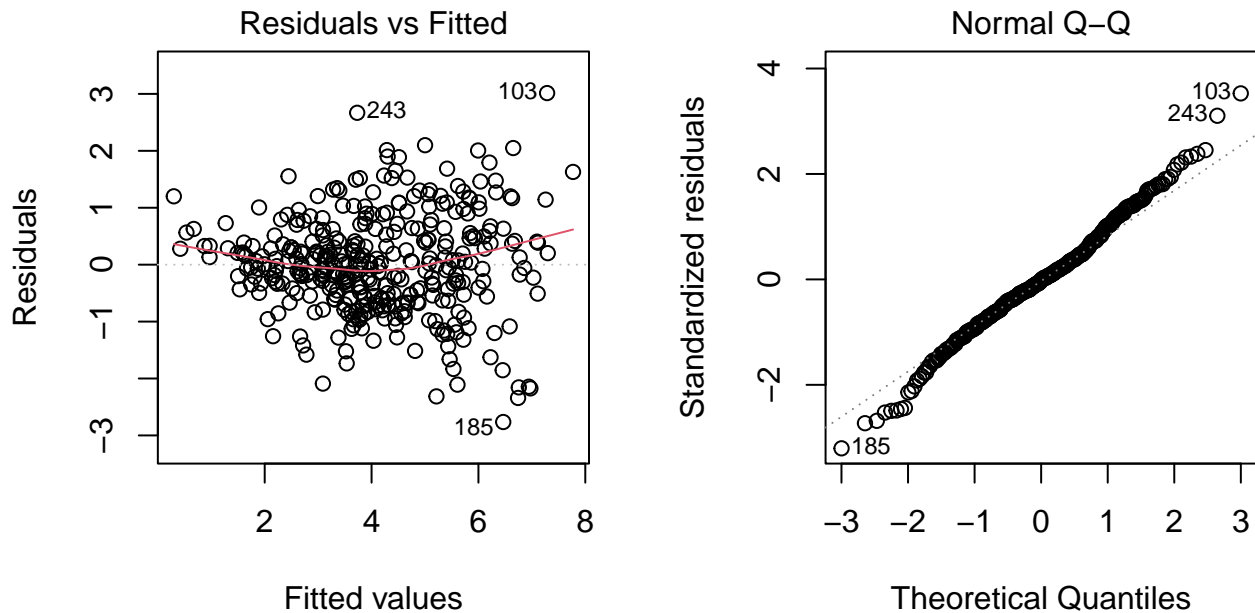
Linear Model Residual and QQ plot

```
par(mfrow = c(1, 2))
plot(models$STROKE_CrudePrev, 1:2)
```



Linear Model after LASSO Residual and QQ plot

```
par(mfrow = c(1, 2))
plot(models_min$STROKE_CrudePrev, 1:2)
```



The linearity assumption for both models is met since the residual plot show reasonable symmetry.

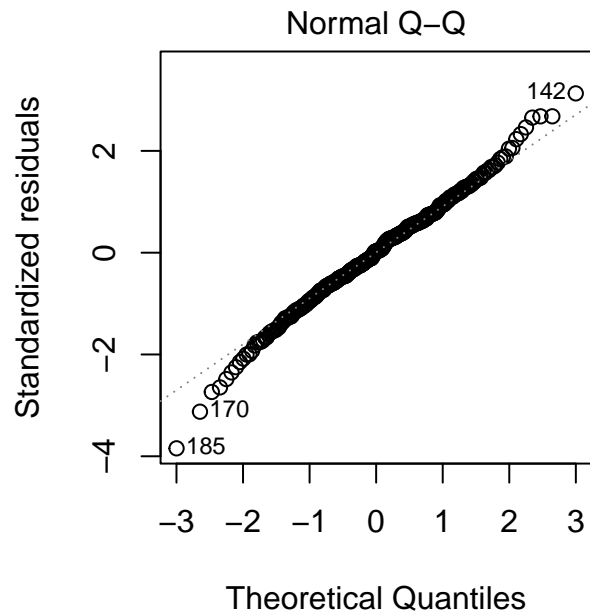
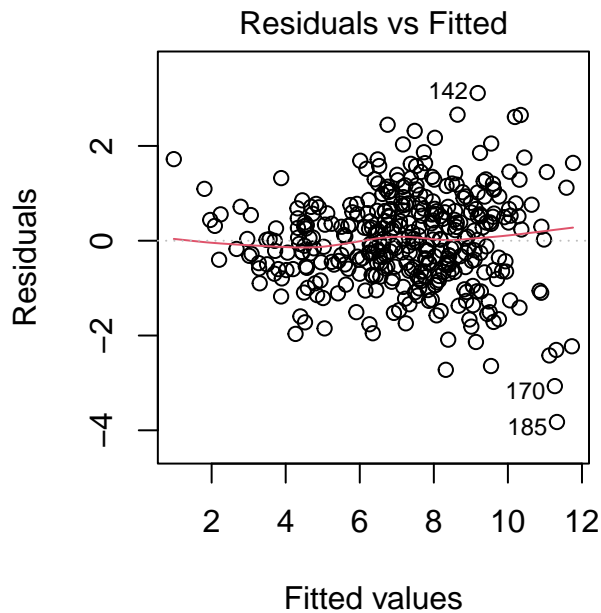
```
anova(models$STROKE_CrudePrev, models_min$STROKE_CrudePrev)
```

```
## Analysis of Variance Table
##
## Model 1: STROKE_CrudePrev ~ Population2010 + HSI_SCORE + HEI_SCORE + HVI_SCORE +
##   sumTOTAL_LPSS + avgLPSS_PER1000 + sumTOTAL_HPSS + avgHPSS_PER1000 +
##   avgPCT_HPSS + avgPCT_VEHICLE_AVAIL + sumTOTAL_RESTAURANTS +
##   avgPCT_POVERTY + sumTOTAL_UNITS + walk_score + transit_score +
##   bike_score + gini + n_hospitals + heat_health_effects
## Model 2: STROKE_CrudePrev ~ Population2010 + HSI_SCORE + HEI_SCORE + sumTOTAL_HPSS +
##   avgHPSS_PER1000 + avgPCT_VEHICLE_AVAIL + sumTOTAL_RESTAURANTS +
##   walk_score + gini + heat_health_effects
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      349 258.86
## 2      358 273.94 -9   -15.087 2.2601 0.01812 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

COPD

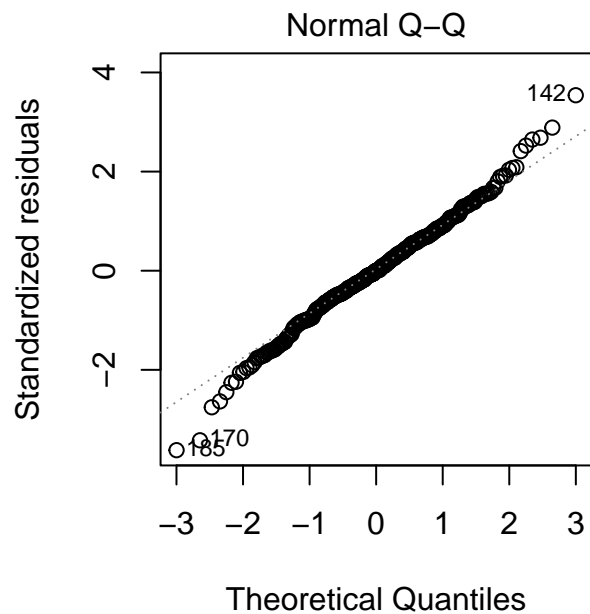
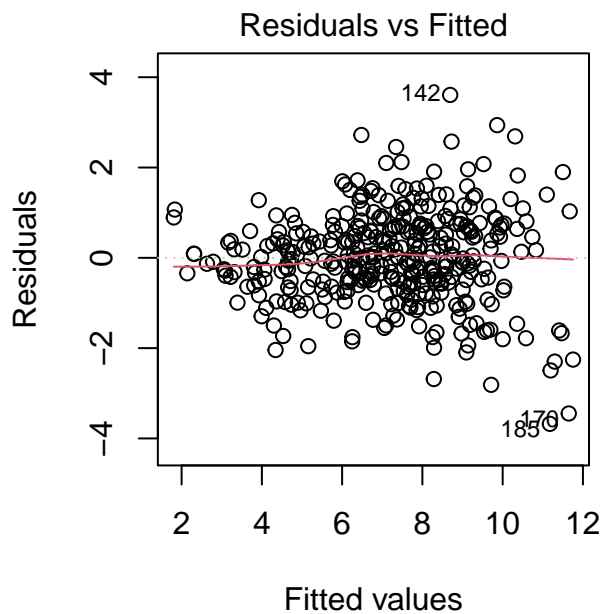
Linear Model Residual and QQ plot

```
par(mfrow = c(1, 2))
plot(models$COPD_CrudePrev, 1:2)
```



Linear Model after LASSO Residual and QQ plot

```
par(mfrow = c(1, 2))
plot(models_min$COPD_CrudePrev, 1:2)
```



```
anova(models$COPD_CrudePrev, models_min$COPD_CrudePrev)
```

```
## Analysis of Variance Table
##
## Model 1: COPD_CrudePrev ~ Population2010 + HSI_SCORE + HEI_SCORE + HVI_SCORE +
##   sumTOTAL_LPSS + avgLPSS_PER1000 + sumTOTAL_HPSS + avgHPSS_PER1000 +
##   avgPCT_HPSS + avgPCT_VEHICLE_AVAIL + sumTOTAL_RESTAURANTS +
##   avgPCT_POVERTY + sumTOTAL_UNITS + walk_score + transit_score +
##   bike_score + gini + n_hospitals + heat_health_effects
```

```
## Model 2: COPD_CrudePrev ~ HEI_SCORE + avgHPSS_PER1000 + avgPCT_HPSS +
##      avgPCT_VEHICLE_AVAIL + sumTOTAL_RESTAURANTS + avgPCT_POVERTY +
##      walk_score + transit_score + gini + heat_health_effects
## Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      349 361.07
## 2      358 379.61 -9      -18.54 1.9911 0.03947 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Appendix I: Data Cleaning

```
## [1] "3352b422b0ef5aaabae7a20651522b5a7688e0bf"
```

Data landing and inspection

In order to directly download data from the Census API, you need a [key](#). You can sign up for a free key here. Type your key in quotes using the `census_api_key()` command.

376 tracts here

384 census tracts here (2010 data)

384 census tracts here (2010 data)

```
## OBJECTID_1 TractFIPS NAME10 OBJECTID HSI_SCORE HEI_SCORE HVI_SCORE
## 1          276 4.2101e+10      1      188 0.9768257 -3.31447 -3.477334
## N_VERYHIGH Shape__Area Shape__Length
## 1          0      1202257      4528.079
```

philly\_census\_health x population\_philly (by race)

```
## [1] 376 77
```

```
## [1] 376 86
```

Next we just need to separate out the last digit in the food\_access GEOID10 to transform the census block group number into a census tract number (<https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html#:~:text=Census%20Tract,482012231001>)

The new census tract column is now `TractFIPS` like the other datasets.

Then we can average the estimates for each census tract to get one row of data for each census tract.

```
## # A tibble: 6 x 9
## TractFIPS sumTOTAL_LPSS avgLPSS_PER1000 sumTOTAL_HPSS avgHPSS_PER1000
## <chr>      <int>      <dbl>      <dbl>      <dbl>
## 1 42101000100      106      25.2      19.5      4.60
## 2 42101000200      120      44.7      17.8      6.62
## 3 42101000300      182      77.7      35.2     14.1
## 4 42101000401       74      28.1       20      7.60
## 5 42101000402      199      59.5      45.8     13.8
## 6 42101000500      148      58.9      23.8     9.45
## # i 4 more variables: avgPCT_HPSS <dbl>, avgPCT_VEHICLE_AVAIL <dbl>,
## # sumTOTAL_RESTAURANTS <int>, avgPCT_POVERTY <dbl>
```

So now that we are down to about 380 census tracts, we can merge with the other data for `merege3`

```
## [1] 376 94
```

Geolocating addresses (for the housing and hospital data) to census tracts

Getting census tract-level gini indices

```
## [1] "GEOID" "state" "county" "tract" "gini"
```

Merge these with the rest of the data matrix

Preparing hospital data

```
##      OBJECTID      HOSPITAL_NAME      STREET_ADDRESS
## 25          1      Aria Health- Frankford Campus 4900 Frankford Avenue
## 22          2      Aria Health- Torresdale Campus 3998 Red Lion Rd
## 23          3 Belmont Center for Comprehensive Treatment 4200 Monument Road
## 36          4      Chestnut Hill Hospital 8835 Germantown Avenue
## 17          5 The Children's Hospital of Philadelphia 3401 Civic Center Blvd
## 6           6      Einstein Medical Center - Philadelphia 1200 West Tabor Road
##      CITY STATE ZIP_CODE PHONE_NUMBER HOSPITAL_TYPE cxy_lon
## 25 Philadelphia PA 19124 215-831-2000 General medical -75.08039
## 22 Philadelphia PA 19114 215-612-4000 General medical -74.98026
## 23 Philadelphia PA 19131 215-877-2000 Behavioral health -75.21645
## 36 Philadelphia PA 19118 215-248-8200 General medical -75.21210
## 17 Philadelphia PA 19104 215-590-1000 General medical -75.19318
## 6 Philadelphia PA 19141 215-456-7890 General medical -75.14383
##      cxy_lat
## 25 40.02035
## 22 40.06777
## 23 39.99804
## 36 40.07856
## 17 39.94810
## 6 40.03806
```

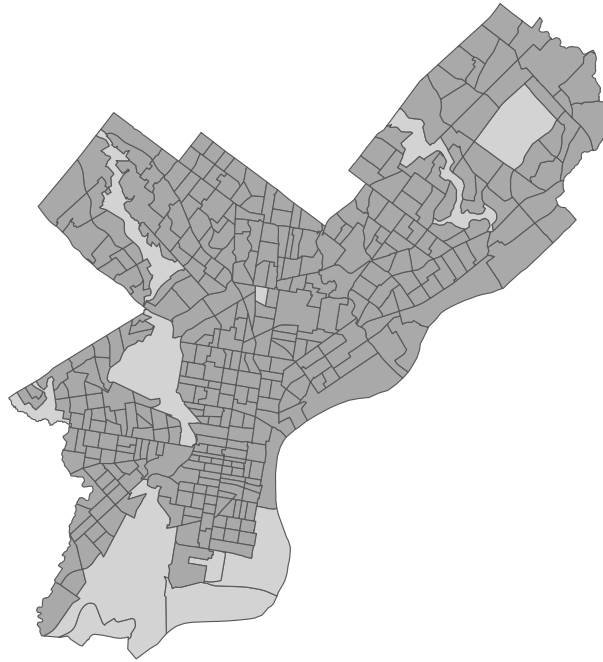
Final merge

## Appendix II: Extended EDA

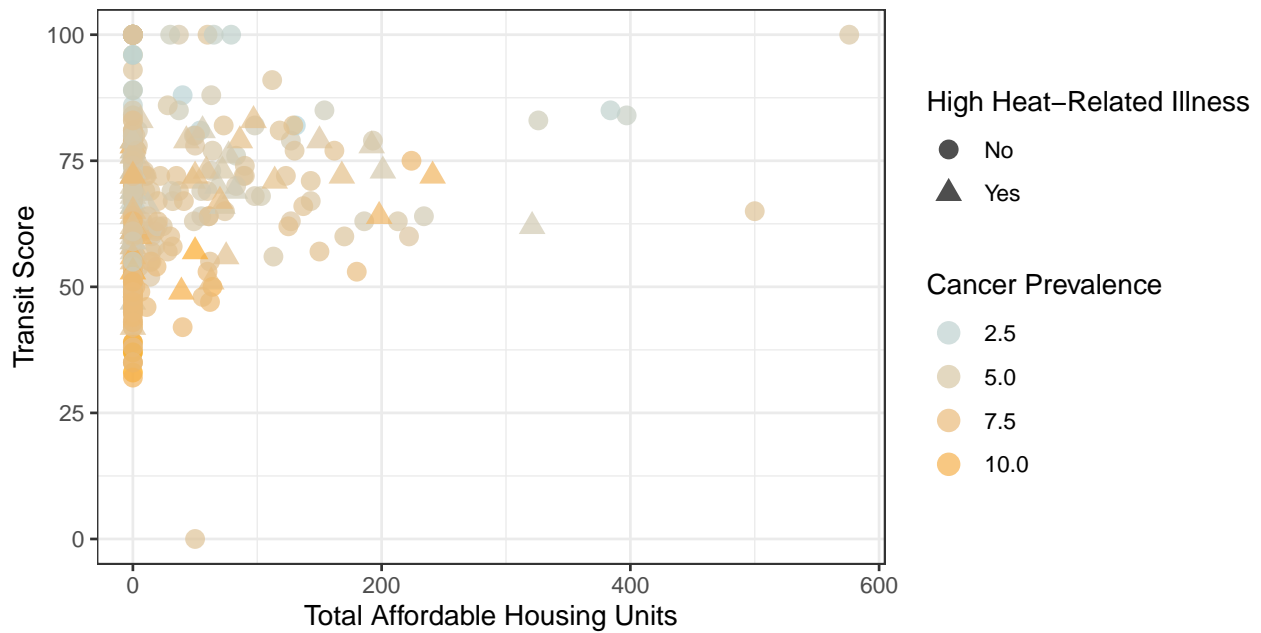
We used only the following residential census tracts for the analysis



## Philadelphia Census Tracts



### Affordable Housing and Transit by Cancer Prevalence and Heat-Related Illness



**EDA looking at relationship of our variables**

[distributions of hospitals]

[distribution of affordable housing]

## Appendix III: Data Dictionary

Here is the break down of variable names:

Variable	Description	Variable	Description
TractFIPS	Census tract identifier	PERCENT_ASIAN_NH	Percent Asian
Population2010	Population in 2010	PERCENT_HISPANIC	Percent Hispanic
ACCESS2_CrudePrev	Access to health insurance prevalence		
ARTHRITIS_CrudePrev	Arthritis prevalence	HSI_SCORE	Displays sensitivity to heat by census tract, incorporating demographic, health and disability indicators (2019)
BINGE_CrudePrev	Binge drinking		
BPHIGH_CrudePrev	High blood pressure	HEI_SCORE	Displays heat exposure by census tract incorporating daytime and nighttime land surface temperature, surface reflectivity, building density, and vegetation (2017-2019).
BPMED_CrudePrev	On blood pressure medication		
CANCER_CrudePrev	cancer	HVI_SCORE	Displays heat vulnerability by census tract, incorporating heat exposure and sensitivity indicators (2017 - 2019).
CASTHMA_CrudePrev	asthma		
CHD_CrudePrev	heart disease	sumTOTAL_LPSS	Total low produce supply stores within a mile of internal census block groups
CHECKUP_CrudePrev	up-to-date on checkups	avgLPSS_PER1000	Average number of low-produce supply stores per 1k people
CHOLSCREEN_CrudePrev	cholesterol screen		
COLON_SCREEN_CrudePrev	colon screen	sumTOTAL_HPSS	Total high produce supply stores within a mile of internal census block groups
COPD_CrudePrev	COPD	avgHPSS_PER1000	Average number of high-produce supply stores per 1k people
COREM_CrudePrev	male core checkups		
COREW_CrudePrev	female core checkups	avgPCT_HPSS	Average percentang of all stores within half mile walking distance of the block group that are high produce supply
CSMOKING_CrudePrev	smoking	sumTOTAL_RESTAURANTS	Total restaurants in the tract
DENTAL_CrudePrev	dental visits	avgPCT_POVERTY	Average percent of people in poverty across census block groups
DIABETES_CrudePrev	diabetes		
HIGHCHOL_CrudePrev	high cholesterol	sumTOTAL_UNITS	Total affordable housing unit within the census tract
KIDNEY_CrudePrev	kidney disease	walk_score	Redfin walk score
LPA_CrudePrev	leisure-time physical activity	transit_score	REdfin transit score
MAMMOUSE_CrudePrev	mamogram use	bike_score	Redfin bike score
MHLTH_CrudePrev	mental health poor for > 14 days	gini	Gini index
OBESITY_CrudePrev	obesity	n_hospitals	Number of hospitals int he census tract
PAPTEST_CrudePrev	pap test		
PHLTH_CrudePrev	poor physical health	heat_health_effects	Indicator variable caputring if this census tract is among the topp 75 most vulnerable to experience health impacts of heat vulnerability
SLEEP_CrudePrev	sleep issues		
STROKE_CrudePrev	stroke		
TEETHLOST_CrudePrev	teeth lost		
COUNT_WHITE_NH	Count of White residents		
COUNT_BLACK_NH	Count of Black residents		
COUNT_ASIAN_NH	Count of Asian residents		
COUNT_HISPANIC	Count of Hispanic Residents		
PERCENT_WHITE_NH	Percent White		
PERCENT_BLACK_NH	Percent Black		
PERCENT_ASIAN_NH	Percent Asian		
PERCENT_HISPANIC	Percent Hispanic		

Figure 2: Data Dictionary