Analysis before fitting the CAR model

Alvin Sheng

6/28/2021

```
library(here)
```

```
## here() starts at /Users/Alvin/Documents/NCSU_Fall_2021/NIH_SIP/flood-risk-health-effects
```

```
library(ape)
library(GGally)
```

```
## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(usdm)
```

```
## Loading required package: sp

## Loading required package: raster

##
## Attaching package: 'raster'

## The following objects are masked from 'package:ape':
##
##     rotate, zoom
```

```
library(spdep)
```

```
## Loading required package: spData

## To access larger datasets in this package, install the spDataLarge
## package with: `install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source')`

## Loading required package: sf

## Linking to GEOS 3.8.1, GDAL 3.1.4, PROJ 6.3.1

## Registered S3 method overwritten by 'spdep':
##   method   from
##   plot.mst ape
```

```
fhs_model_df <- readRDS(here("intermediary_data/fhs_model_df_all_census_tract_reorg.rds"))
```

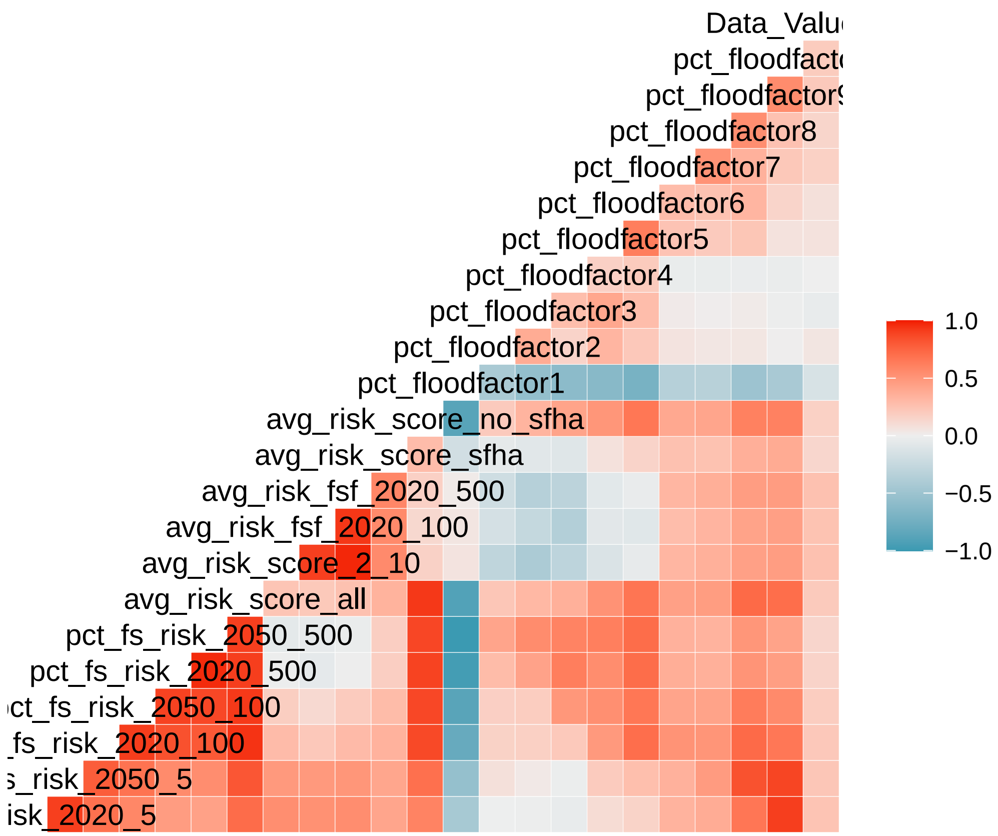
Checking for multicollinearity among the covariates

`S.CARleroux()` automatically puts a fixed ridge penalty on the beta coefficients. Therefore, the large number of covariates and multicollinearity would be accounted for.

Actually no, because the penalty is negligible.

Flood risk variables

```
ggcorr(data = fhs_model_df[, c(14:35, ncol(fhs_model_df))])
```




```
flood_cor <- cor(fhs_model_df[complete.cases(fhs_model_df[, c(14:35, ncol(fhs_model_df))]), c(14:35, nc

flood_cor[nrow(flood_cor), ] # correlation with dependent variable
```

```
##     pct_fs_risk_2020_5     pct_fs_risk_2050_5   pct_fs_risk_2020_100
##            0.242152196            0.229647057            0.219185199
##   pct_fs_risk_2050_100   pct_fs_risk_2020_500   pct_fs_risk_2050_500
##            0.193017678            0.157476172            0.146254655
##     avg_risk_score_all    avg_risk_score_2_10   avg_risk_fsf_2020_100
##            0.210887927            0.258108501            0.245562831
##  avg_risk_fsf_2020_500    avg_risk_score_sfha avg_risk_score_no_sfha
##            0.258140879            0.136061159            0.173556262
##       pct_floodfactor1       pct_floodfactor2       pct_floodfactor3
##           -0.146301786            0.055788581           -0.033703567
##       pct_floodfactor4       pct_floodfactor5       pct_floodfactor6
##            0.004699415            0.071103011            0.089231160
```

```
##       pct_floodfactor7       pct_floodfactor8       pct_floodfactor9
##            0.177730956            0.145830985            0.211774217
##      pct_floodfactor10          Data_Value_CHD
##            0.198914543            1.000000000
```

For each variable, I take the summary of its correlations with other variables, not including itself.

```
diag(flood_cor) <- NA

summary(flood_cor)
```

```
##  pct_fs_risk_2020_5 pct_fs_risk_2050_5 pct_fs_risk_2020_100
##  Min.   :-0.4369    Min.   :-0.5491    Min.   :-0.8081
##  1st Qu.: 0.1850    1st Qu.: 0.2446    1st Qu.: 0.2337
##  Median : 0.4502    Median : 0.4807    Median : 0.5224
##  Mean   : 0.3984    Mean   : 0.4423    Mean   : 0.4815
##  3rd Qu.: 0.5956    3rd Qu.: 0.6966    3rd Qu.: 0.7758
##  Max.   : 0.9094    Max.   : 0.9019    Max.   : 0.9443
##  NA's   :1          NA's   :1          NA's   :1
##  pct_fs_risk_2050_100 pct_fs_risk_2020_500 pct_fs_risk_2050_500
##  Min.   :-0.8792      Min.   :-0.9662      Min.   :-1.0000
##  1st Qu.: 0.1947      1st Qu.: 0.2104      1st Qu.: 0.2214
##  Median : 0.5237      Median : 0.4603      Median : 0.4721
##  Mean   : 0.4536      Mean   : 0.4151      Mean   : 0.4154
##  3rd Qu.: 0.6806      3rd Qu.: 0.6984      3rd Qu.: 0.6987
##  Max.   : 0.9269      Max.   : 0.9661      Max.   : 0.9661
##  NA's   :1            NA's   :1            NA's   :1
##  avg_risk_score_all avg_risk_score_2_10 avg_risk_fsf_2020_100
##  Min.   :-0.9074    Min.   :-0.41684    Min.   :-0.37456
##  1st Qu.: 0.2654    1st Qu.:-0.04624    1st Qu.:-0.07014
##  Median : 0.4974    Median : 0.24126    Median : 0.20744
##  Mean   : 0.4978    Mean   : 0.22193    Mean   : 0.21475
##  3rd Qu.: 0.7952    3rd Qu.: 0.45699    3rd Qu.: 0.44623
##  Max.   : 0.9443    Max.   : 0.97345    Max.   : 0.93399
##  NA's   :1          NA's   :1           NA's   :1
##  avg_risk_fsf_2020_500 avg_risk_score_sfha avg_risk_score_no_sfha
##  Min.   :-0.35201      Min.   :-0.1812     Min.   :-0.8778
##  1st Qu.:-0.03361      1st Qu.: 0.1408     1st Qu.: 0.2232
##  Median : 0.25325      Median : 0.2711     Median : 0.4612
##  Mean   : 0.23936      Mean   : 0.2438     Mean   : 0.4529
##  3rd Qu.: 0.46540      3rd Qu.: 0.3802     3rd Qu.: 0.6919
##  Max.   : 0.97345      Max.   : 0.5814     Max.   : 0.9287
##  NA's   :1             NA's   :1           NA's   :1
##  pct_floodfactor1   pct_floodfactor2    pct_floodfactor3   pct_floodfactor4
##  Min.   :-0.99998   Min.   :-0.423769   Min.   :-0.55620   Min.   :-0.599291
##  1st Qu.:-0.78662   1st Qu.:-0.002819   1st Qu.:-0.02788   1st Qu.:-0.036417
##  Median :-0.52805   Median : 0.077338   Median : 0.03219   Median :-0.006164
##  Mean   :-0.50617   Mean   : 0.077974   Mean   : 0.07911   Mean   : 0.075734
##  3rd Qu.:-0.34602   3rd Qu.: 0.220125   3rd Qu.: 0.31086   3rd Qu.: 0.259810
##  Max.   : 0.07821   Max.   : 0.423576   Max.   : 0.55560   Max.   : 0.626455
##  NA's   :1          NA's   :1           NA's   :1          NA's   :1
##  pct_floodfactor5   pct_floodfactor6   pct_floodfactor7  pct_floodfactor8
##  Min.   :-0.62725   Min.   :-0.7223   Min.   :-0.3616   Min.   :-0.3408
##  1st Qu.: 0.07445   1st Qu.: 0.1563   1st Qu.: 0.2437   1st Qu.: 0.2171
##  Median : 0.22203   Median : 0.2796   Median : 0.3284   Median : 0.3442
```

```
##  Mean   : 0.23104   Mean   : 0.2924   Mean   : 0.2799   Mean   : 0.2873
##  3rd Qu.: 0.49750   3rd Qu.: 0.6592   3rd Qu.: 0.3826   3rd Qu.: 0.4276
##  Max.   : 0.64319   Max.   : 0.7225   Max.   : 0.5569   Max.   : 0.5569
##  NA's   :1          NA's   :1         NA's   :1         NA's   :1
##  pct_floodfactor9  pct_floodfactor10 Data_Value_CHD
##  Min.   :-0.5070   Min.   :-0.4299   Min.   :-0.1463
##  1st Qu.: 0.2646   1st Qu.: 0.1697   1st Qu.: 0.1009
##  Median : 0.4446   Median : 0.4416   Median : 0.1756
##  Mean   : 0.3971   Mean   : 0.3648   Mean   : 0.1475
##  3rd Qu.: 0.6047   3rd Qu.: 0.5605   3rd Qu.: 0.2173
##  Max.   : 0.8375   Max.   : 0.9094   Max.   : 0.2581
##  NA's   :1         NA's   :1         NA's   :1
```

Many of the flood risk variables are very correlated.

Using VIF to exlude variables

```
fhs_model_df <- readRDS(here("intermediary_data/fhs_model_df_all_census_tract_reorg.rds"))
```

```
X <- fhs_model_df[, 14:(ncol(fhs_model_df) - 1)]

X <- X[, names(X) != "pct_floodfactor1"]

X <- X[, names(X) != "avg_risk_score_sfha"]



X            <- scale(X) # Scale covariates

X <- data.frame(X)
```

```
vif(X)
```

```
##                  Variables          VIF
## 1      pct_fs_risk_2020_5 8.771527e+00
## 2      pct_fs_risk_2050_5 2.786375e+01
## 3    pct_fs_risk_2020_100 3.630384e+01
## 4    pct_fs_risk_2050_100 2.428735e+01
## 5    pct_fs_risk_2020_500 4.668786e+01
## 6    pct_fs_risk_2050_500 1.039171e+04
## 7      avg_risk_score_all 1.222366e+05
## 8     avg_risk_score_2_10 4.268659e+01
## 9   avg_risk_fsf_2020_100 1.108746e+01
## 10  avg_risk_fsf_2020_500 4.622057e+01
## 11 avg_risk_score_no_sfha 1.079307e+01
## 12       pct_floodfactor2 2.443883e+02
## 13       pct_floodfactor3 1.847160e+03
## 14       pct_floodfactor4 1.072501e+04
## 15       pct_floodfactor5 1.144970e+03
## 16       pct_floodfactor6 1.469482e+04
## 17       pct_floodfactor7 1.501687e+03
## 18       pct_floodfactor8 1.650923e+02
## 19       pct_floodfactor9 1.203249e+04
```

```
## 20      pct_floodfactor10 2.225740e+04
## 21                 EP_POV 3.864096e+00
## 22               EP_UNEMP 1.909525e+00
## 23                 EP_PCI 2.813493e+00
## 24              EP_NOHSDP 5.883226e+00
## 25               EP_AGE65 2.377762e+00
## 26               EP_AGE17 2.899834e+00
## 27              EP_DISABL 2.764914e+00
## 28              EP_SNGPNT 2.733916e+00
## 29              EP_MINRTY 3.796610e+00
## 30              EP_LIMENG 4.406225e+00
## 31               EP_MUNIT 2.090833e+00
## 32              EP_MOBILE 1.612867e+00
## 33               EP_CROWD 2.731768e+00
## 34               EP_NOVEH 3.392003e+00
## 35              EP_GROUPQ 1.528858e+00
## 36             EP_UNINSUR 2.518945e+00
## 37                     co 1.061800e+01
## 38                    no2 1.641811e+01
## 39                     o3 2.844783e+00
## 40                   pm10 4.145901e+00
## 41                   pm25 5.149004e+00
## 42                    so2 2.853547e+00
## 43            summer_tmmx 4.549515e+00
## 44            winter_tmmx 5.326834e+00
## 45            summer_rmax 4.177491e+00
## 46            winter_rmax 3.632492e+00
## 47    Data_Value_CSMOKING 5.933009e+00
```

```
vifstep(X)
```

```
## 8 variables from the 47 input variables have collinearity problem:
##
## avg_risk_score_all pct_fs_risk_2050_500 pct_fs_risk_2020_500 avg_risk_fsf_2020_500 pct_fs_risk_2050_5
##
## After excluding the collinear variables, the linear correlation coefficients ranges between:
## min correlation ( so2 ~ EP_LIMENG ):  7.815658e-05
## max correlation ( avg_risk_fsf_2020_100 ~ avg_risk_score_2_10 ):  0.9049275
##
## ---------- VIFs of the remained variables --------
##                Variables      VIF
## 1      pct_fs_risk_2020_5 7.483365
## 2     avg_risk_score_2_10 9.846115
## 3   avg_risk_fsf_2020_100 9.094658
## 4  avg_risk_score_no_sfha 9.308518
## 5        pct_floodfactor2 1.465039
## 6        pct_floodfactor3 1.814429
## 7        pct_floodfactor4 2.534454
## 8        pct_floodfactor5 1.940262
## 9        pct_floodfactor6 3.255760
## 10       pct_floodfactor7 2.031480
## 11       pct_floodfactor8 2.012140
## 12       pct_floodfactor9 3.107951
## 13      pct_floodfactor10 7.469637
## 14                 EP_POV 3.706739
```

```
## 15                EP_UNEMP 1.859158
## 16                  EP_PCI 2.816565
## 17               EP_NOHSDP 6.008866
## 18                EP_AGE65 2.406046
## 19                EP_AGE17 2.859408
## 20               EP_DISABL 2.744534
## 21               EP_SNGPNT 2.782196
## 22               EP_MINRTY 3.917627
## 23               EP_LIMENG 4.261899
## 24                EP_MUNIT 2.057475
## 25               EP_MOBILE 1.584882
## 26                EP_CROWD 3.091777
## 27                EP_NOVEH 3.209118
## 28               EP_GROUPQ 1.375356
## 29              EP_UNINSUR 2.602599
## 30                      co 5.347401
## 31                      o3 2.731749
## 32                    pm10 4.001244
## 33                    pm25 4.255885
## 34                     so2 2.907334
## 35             summer_tmmx 4.480544
## 36             winter_tmmx 4.650053
## 37             summer_rmax 3.828017
## 38             winter_rmax 3.221183
## 39     Data_Value_CSMOKING 5.713966
```

This procedure detects that the following variables have collinearity problems. Let's exclude these variables and then rerun the analysis.

```
collin_var_names <- c("avg_risk_score_all", "pct_fs_risk_2050_500", "pct_fs_risk_2020_500", "avg_risk_fs
```

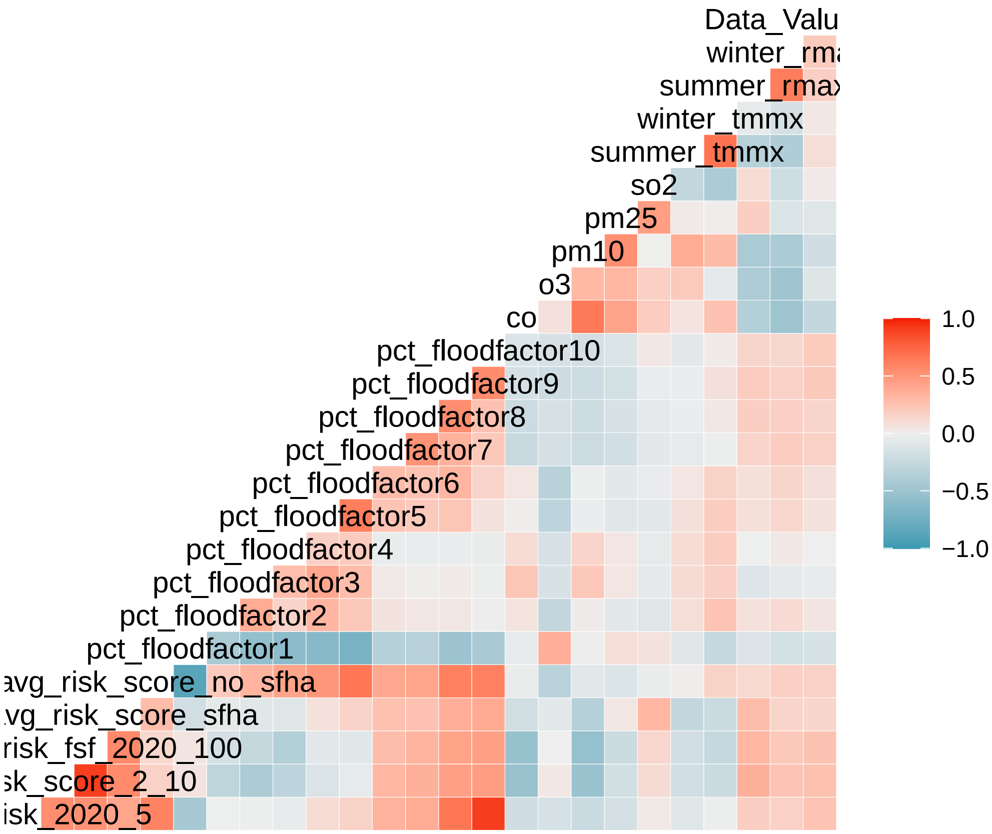
Correlations among climate related variables

Excluding variables in collin_var_names

```
climate_var_idx <- c(14:35, 52:61)

climate_var_idx_exclude <- climate_var_idx[-which(names(fhs_model_df)[climate_var_idx] %in% collin_var_n
```

```
ggcorr(data = fhs_model_df[, c(climate_var_idx_exclude, ncol(fhs_model_df))])
```




```r
climate_cor <- cor(fhs_model_df[complete.cases(fhs_model_df[, c(climate_var_idx_exclude, ncol(fhs_model_

climate_cor[nrow(climate_cor), ] # correlation with dependent variable
```

```
##     pct_fs_risk_2020_5    avg_risk_score_2_10  avg_risk_fsf_2020_100
##            0.245316056            0.258075405            0.245475391
##    avg_risk_score_sfha avg_risk_score_no_sfha       pct_floodfactor1
##            0.135898990            0.173772974           -0.146150862
##       pct_floodfactor2       pct_floodfactor3       pct_floodfactor4
##            0.055842150           -0.033632706            0.004753925
##       pct_floodfactor5       pct_floodfactor6       pct_floodfactor7
##            0.071181020            0.089300494            0.178169288
##       pct_floodfactor8       pct_floodfactor9      pct_floodfactor10
##            0.146041633            0.211877045            0.201828766
##                     co                     o3                   pm10
##           -0.265707019           -0.110967465           -0.177872621
##                   pm25                    so2            summer_tmmx
##           -0.081799241            0.035606141            0.093375275
##            winter_tmmx            summer_rmax            winter_rmax
##            0.050894297            0.184535350            0.198197705
##         Data_Value_CHD
##            1.000000000
```

For each variable, I take the summary of its correlations with other variables, not including itself.

```r
diag(climate_cor) <- NA

summary(climate_cor)
```

```
##  pct_fs_risk_2020_5 avg_risk_score_2_10 avg_risk_fsf_2020_100
##  Min.   :-0.43306   Min.   :-0.51695    Min.   :-0.52800
##  1st Qu.:-0.05438   1st Qu.:-0.21132    1st Qu.:-0.18754
##  Median : 0.13299   Median : 0.11398    Median : 0.09563
##  Mean   : 0.16658   Mean   : 0.08348    Mean   : 0.08019
##  3rd Qu.: 0.38957   3rd Qu.: 0.34574    3rd Qu.: 0.31581
##  Max.   : 0.90603   Max.   : 0.90647    Max.   : 0.90647
##  NA's   :1          NA's   :1           NA's   :1
##  avg_risk_score_sfha avg_risk_score_no_sfha pct_floodfactor1
##  Min.   :-0.3421     Min.   :-0.87775       Min.   :-0.87775
##  1st Qu.:-0.0820     1st Qu.: 0.00214       1st Qu.:-0.45094
##  Median : 0.1391     Median : 0.17764       Median :-0.22006
##  Mean   : 0.1142     Mean   : 0.18999       Mean   :-0.25848
##  3rd Qu.: 0.2920     3rd Qu.: 0.42014       3rd Qu.:-0.04719
##  Max.   : 0.5608     Max.   : 0.66698       Max.   : 0.38204
##  NA's   :1           NA's   :1              NA's   :1
##  pct_floodfactor2   pct_floodfactor3   pct_floodfactor4    pct_floodfactor5
##  Min.   :-0.42518   Min.   :-0.55819   Min.   :-0.601194   Min.   :-0.62923
##  1st Qu.:-0.05919   1st Qu.:-0.05729   1st Qu.:-0.040417   1st Qu.:-0.02835
##  Median : 0.05257   Median : 0.01794   Median : 0.001479   Median : 0.08506
##  Mean   : 0.03089   Mean   : 0.03456   Mean   : 0.005256   Mean   : 0.09500
##  3rd Qu.: 0.12780   3rd Qu.: 0.22528   3rd Qu.: 0.148384   3rd Qu.: 0.21365
##  Max.   : 0.37421   Max.   : 0.41215   Max.   : 0.423418   Max.   : 0.64319
##  NA's   :1          NA's   :1          NA's   :1           NA's   :1
##  pct_floodfactor6  pct_floodfactor7   pct_floodfactor8   pct_floodfactor9
##  Min.   :-0.7242   Min.   :-0.35823   Min.   :-0.33938   Min.   :-0.50458
##  1st Qu.:-0.0264   1st Qu.:-0.05244   1st Qu.:-0.03318   1st Qu.:-0.02355
##  Median : 0.1595   Median : 0.16993   Median : 0.15566   Median : 0.17411
##  Mean   : 0.1135   Mean   : 0.11130   Mean   : 0.12555   Mean   : 0.16609
##  3rd Qu.: 0.2370   3rd Qu.: 0.29409   3rd Qu.: 0.28282   3rd Qu.: 0.38561
##  Max.   : 0.6670   Max.   : 0.55584   Max.   : 0.55584   Max.   : 0.65822
##  NA's   :1         NA's   :1          NA's   :1          NA's   :1
##  pct_floodfactor10        co                 o3                pm10
##  Min.   :-0.42600   Min.   :-0.51695   Min.   :-0.49726   Min.   :-0.52800
##  1st Qu.:-0.03784   1st Qu.:-0.21098   1st Qu.:-0.22845   1st Qu.:-0.22379
##  Median : 0.09872   Median :-0.03639   Median :-0.14578   Median :-0.04783
##  Mean   : 0.14908   Mean   :-0.04806   Mean   :-0.08411   Mean   :-0.04226
##  3rd Qu.: 0.28952   3rd Qu.: 0.08641   3rd Qu.: 0.05269   3rd Qu.: 0.16407
##  Max.   : 0.90603   Max.   : 0.61902   Max.   : 0.38204   Max.   : 0.61902
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##       pm25               so2            summer_tmmx        winter_tmmx
##  Min.   :-0.18009   Min.   :-0.40848   Min.   :-0.38951   Min.   :-0.40848
##  1st Qu.:-0.12574   1st Qu.:-0.05044   1st Qu.:-0.12996   1st Qu.:-0.10573
##  Median :-0.07480   Median :-0.01482   Median :-0.01156   Median : 0.03016
##  Mean   : 0.02257   Mean   : 0.02038   Mean   :-0.00538   Mean   : 0.03649
##  3rd Qu.: 0.05742   3rd Qu.: 0.12728   3rd Qu.: 0.09359   3rd Qu.: 0.18067
##  Max.   : 0.50234   Max.   : 0.47229   Max.   : 0.69477   Max.   : 0.69477
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##   summer_rmax        winter_rmax       Data_Value_CHD
##  Min.   :-0.41426   Min.   :-0.49726   Min.   :-0.265707
##  1st Qu.:-0.05547   1st Qu.:-0.16132   1st Qu.:-0.004843
##  Median : 0.12376   Median : 0.12718   Median : 0.091338
##  Mean   : 0.06470   Mean   : 0.01933   Mean   : 0.073500
##  3rd Qu.: 0.18600   3rd Qu.: 0.16934   3rd Qu.: 0.187951
```

```
## Max.    : 0.62570   Max.    : 0.62570   Max.    : 0.258075
## NA's    :1          NA's    :1          NA's    :1
```

Climate variables other than flood risk are not too correlated.

Non-spatial modeling

```
Y <- fhs_model_df$Data_Value_CHD

X <- fhs_model_df[, 14:(ncol(fhs_model_df) - 1)]

X <- X[, names(X) != "pct_floodfactor1"]

# exclude some more variables selected by vifstep, to account for multicollinearity
# excluding all of the pct_fs_risk variables, as well as 3 of the avg_risk_score variables

collin_var_names <- c("avg_risk_score_all", "pct_fs_risk_2050_500", "pct_fs_risk_2020_500", "avg_risk_f

X <- X[, !(names(X) %in% collin_var_names)]

# also removing avg_risk_score_sfha due to large numbers of NAs
X <- X[, names(X) != "avg_risk_score_sfha"]



X            <- scale(X) # Scale covariates
X[is.na(X)] <- 0         # Fill in missing values with the mean

# if I do mean imputation (which may be problematic), all the counties
# will have neighbors in W

# X <- data.frame(X)
```

```
fhs_lm <- lm(Y ~ X)
```

```
summary(fhs_lm)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.9654 -0.4812 -0.0178  0.4570 17.8394
##
## Coefficients:
##                           Estimate Std. Error  t value Pr(>|t|)
## (Intercept)               6.664936   0.003135 2126.210  < 2e-16 ***
## Xpct_fs_risk_2020_5       0.053748   0.009252    5.809 6.29e-09 ***
## Xavg_risk_score_2_10      0.007900   0.009586    0.824 0.409912
## Xavg_risk_fsf_2020_100    0.032057   0.008780    3.651 0.000261 ***
## Xavg_risk_score_no_sfha -0.011752   0.008623   -1.363 0.172961
## Xpct_floodfactor2        -0.004546   0.003816   -1.191 0.233507
```

```
## Xpct_floodfactor3       -0.004014   0.004290   -0.936 0.349446
## Xpct_floodfactor4        0.002632   0.004618    0.570 0.568763
## Xpct_floodfactor5        0.002064   0.004464    0.462 0.643840
## Xpct_floodfactor6        0.003643   0.005293    0.688 0.491284
## Xpct_floodfactor7        0.021413   0.004042    5.297 1.18e-07 ***
## Xpct_floodfactor8       -0.017865   0.004226   -4.227 2.37e-05 ***
## Xpct_floodfactor9       -0.022564   0.005283   -4.271 1.95e-05 ***
## Xpct_floodfactor10      -0.010415   0.009034   -1.153 0.248975
## XEP_POV                  0.338679   0.005919   57.222  < 2e-16 ***
## XEP_UNEMP                0.015666   0.004283    3.658 0.000255 ***
## XEP_PCI                 -0.025324   0.005249   -4.824 1.41e-06 ***
## XEP_NOHSDP               0.210564   0.007408   28.423  < 2e-16 ***
## XEP_AGE65                1.463340   0.004761  307.362  < 2e-16 ***
## XEP_AGE17                0.325402   0.005214   62.414  < 2e-16 ***
## XEP_DISABL               0.343750   0.005125   67.068  < 2e-16 ***
## XEP_SNGPNT              -0.101442   0.005052  -20.080  < 2e-16 ***
## XEP_MINRTY              -0.060975   0.006023  -10.124  < 2e-16 ***
## XEP_LIMENG              -0.007271   0.006244   -1.165 0.244206
## XEP_MUNIT               -0.059793   0.004493  -13.309  < 2e-16 ***
## XEP_MOBILE               0.042921   0.003954   10.855  < 2e-16 ***
## XEP_CROWD               -0.067508   0.005296  -12.746  < 2e-16 ***
## XEP_NOVEH                0.038119   0.005599    6.809 9.94e-12 ***
## XEP_GROUPQ              -0.075081   0.003835  -19.579  < 2e-16 ***
## XEP_UNINSUR              0.150083   0.004833   31.056  < 2e-16 ***
## Xco                      0.021276   0.007194    2.958 0.003101 **
## Xo3                     -0.066440   0.005127  -12.959  < 2e-16 ***
## Xpm10                   -0.004519   0.006259   -0.722 0.470332
## Xpm25                   -0.003595   0.006463   -0.556 0.578044
## Xso2                     0.074978   0.005215   14.377  < 2e-16 ***
## Xsummer_tmmx             0.114205   0.006603   17.295  < 2e-16 ***
## Xwinter_tmmx             0.060476   0.006672    9.064  < 2e-16 ***
## Xsummer_rmax             0.053972   0.006337    8.517  < 2e-16 ***
## Xwinter_rmax             0.082370   0.005649   14.582  < 2e-16 ***
## XData_Value_CSMOKING     0.838028   0.007605  110.194  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8401 on 71785 degrees of freedom
## Multiple R-squared:  0.8552, Adjusted R-squared:  0.8551
## F-statistic: 1.087e+04 on 39 and 71785 DF,  p-value: < 2.2e-16
```

Checking for spatial autocorrelation

```
W <- readRDS(here("intermediary_data", "census_tract_adj_reorganize_all_census_tract.rds"))

W_listw <- mat2listw(W)
```

Moran's I

```
(moran_results <- moran.test(residuals(fhs_lm), W_listw))
```

```
##
##  Moran I test under randomisation
```

```
##
## data:  residuals(fhs_lm)
## weights: W_listw
##
## Moran I statistic standard deviate = 131.76, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      2.789298e-01     -1.392292e-05      4.481720e-06
```

The p-value is negligible, so we can reject the null hypothesis of zero spatial autocorrelation. Since the observed value of I is significantly greater then the expected value, the life expectancies are positively autocorrelated, in contrast to negatively autocorrelated. Thus, using a CAR model is justified.