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
```

```
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
fhs_model_df <- readRDS(here("intermediary_data/fhs_model_df_all_census_tract_reorg.rds"))
```
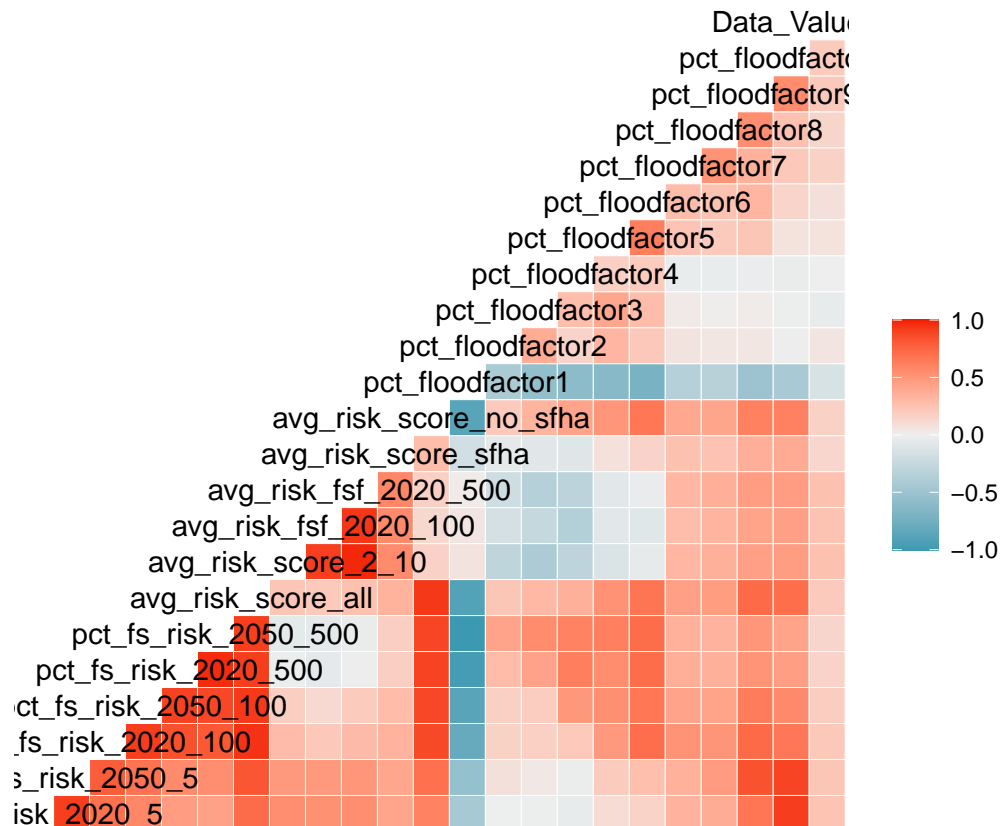
Checking for multicollinearity among the covariates

`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 1.419870e+01
## 2       pct_fs_risk_2050_5 3.106677e+01
## 3     pct_fs_risk_2020_100 3.099006e+01
## 4     pct_fs_risk_2050_100 2.564816e+01
## 5     pct_fs_risk_2020_500 5.102555e+01
## 6     pct_fs_risk_2050_500 5.312171e+04
## 7        avg_risk_score_all 1.197504e+05
## 8       avg_risk_score_2_10 3.980493e+01
## 9    avg_risk_fsf_2020_100 1.062798e+01
## 10   avg_risk_fsf_2020_500 4.395216e+01
## 11 avg_risk_score_no_sfha 9.471106e+00
## 12        pct_floodfactor2 1.532910e+03
## 13        pct_floodfactor3 4.246736e+03
## 14        pct_floodfactor4 1.918893e+04
## 15        pct_floodfactor5 1.695977e+03
## 16        pct_floodfactor6 1.744670e+04
## 17        pct_floodfactor7 2.987486e+03
## 18        pct_floodfactor8 2.296610e+02
## 19        pct_floodfactor9 1.095744e+04
```

```
## 20      pct_floodfactor10 2.669320e+04
## 21                 EP_POV 3.649200e+00
## 22               EP_UNEMP 1.930953e+00
## 23                 EP_PCI 2.906545e+00
## 24              EP_NOHSDP 5.737295e+00
## 25               EP_AGE65 2.462000e+00
## 26               EP_AGE17 2.899173e+00
## 27               EP_DISABL 2.780640e+00
## 28               EP_SNGPNT 2.759617e+00
## 29               EP_MINRTY 3.810426e+00
## 30               EP_LIMENG 4.145805e+00
## 31                EP_MUNIT 2.079027e+00
## 32               EP_MOBILE 1.646938e+00
## 33                EP_CROWD 3.000713e+00
## 34                EP_NOVEH 3.396133e+00
## 35               EP_GROUPQ 1.511203e+00
## 36              EP_UNINSUR 2.499655e+00
## 37                     co 9.517134e+00
## 38                    no2 1.530753e+01
## 39                     o3 2.717140e+00
## 40                   pm10 4.261081e+00
## 41                   pm25 5.314979e+00
## 42                    so2 2.999749e+00
## 43            summer_tmmx 4.187919e+00
## 44            winter_tmmx 5.155049e+00
## 45            summer_rmax 4.029161e+00
## 46            winter_rmax 3.432986e+00
## 47     Data_Value_CSMOKING 6.194212e+00
```

```
vifstep(X)
```

```
## 10 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_!
##
## After excluding the collinear variables, the linear correlation coefficients ranges between:
## min correlation ( EP_GROUPQ ~ pct_floodfactor10 ):  0.0001786664
## max correlation ( avg_risk_fsf_2020_100 ~ avg_risk_score_2_10 ):  0.9029307
##
## ---------- VIFs of the remained variables --------
##               Variables      VIF
## 1    avg_risk_score_2_10 9.983203
## 2  avg_risk_fsf_2020_100 8.534814
## 3       pct_floodfactor2 1.553013
## 4       pct_floodfactor3 1.819917
## 5       pct_floodfactor4 1.402235
## 6       pct_floodfactor5 2.268403
## 7       pct_floodfactor6 2.332049
## 8       pct_floodfactor7 1.659289
## 9       pct_floodfactor8 1.796106
## 10      pct_floodfactor9 2.058287
## 11     pct_floodfactor10 1.750783
## 12               EP_POV 3.536852
## 13             EP_UNEMP 1.856489
## 14               EP_PCI 2.914059
```

```
## 15            EP_NOHSDP 5.836024
## 16             EP_AGE65 2.243527
## 17             EP_AGE17 2.755638
## 18            EP_DISABL 2.730123
## 19            EP_SNGPNT 2.539688
## 20            EP_MINRTY 3.830491
## 21            EP_LIMENG 3.988138
## 22             EP_MUNIT 2.047667
## 23            EP_MOBILE 1.589947
## 24             EP_CROWD 2.905911
## 25             EP_NOVEH 3.189684
## 26            EP_GROUPQ 1.556573
## 27           EP_UNINSUR 2.327937
## 28                   co 5.340469
## 29                   o3 2.701659
## 30                 pm10 4.139363
## 31                 pm25 4.405691
## 32                  so2 2.869026
## 33          summer_tmmx 4.235169
## 34          winter_tmmx 4.623028
## 35          summer_rmax 3.818462
## 36          winter_rmax 3.093104
## 37   Data_Value_CSMOKING 6.126767
```

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_f
```
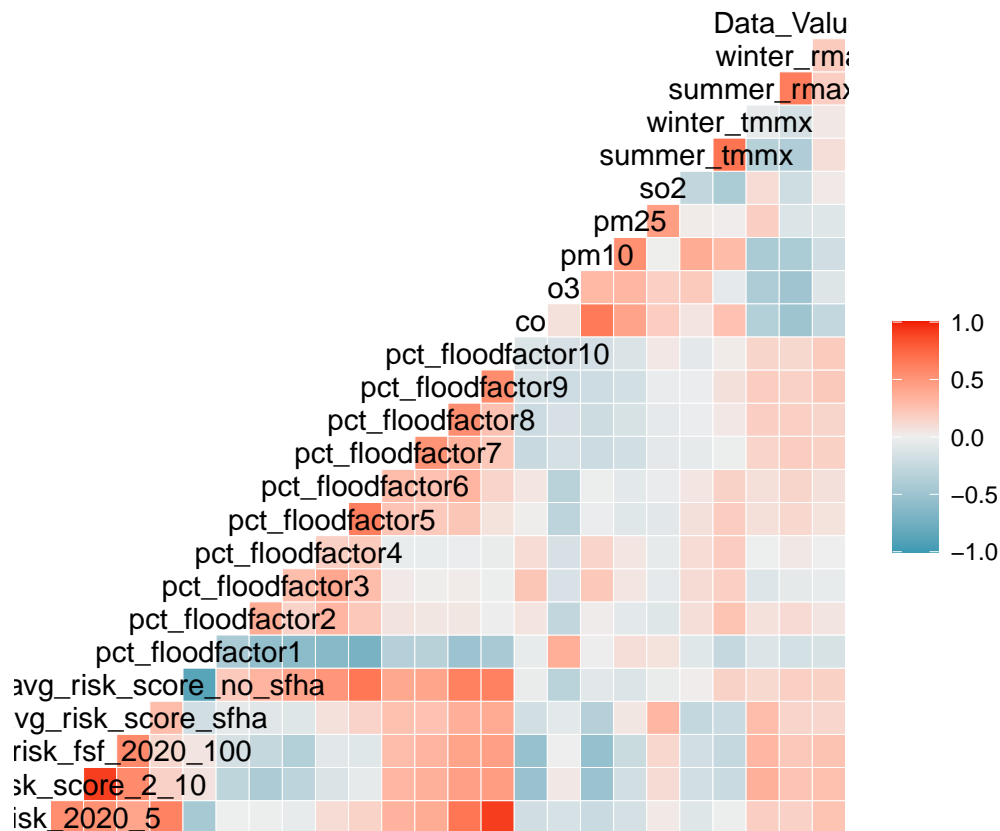
Correlations among climate related variables

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_

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.
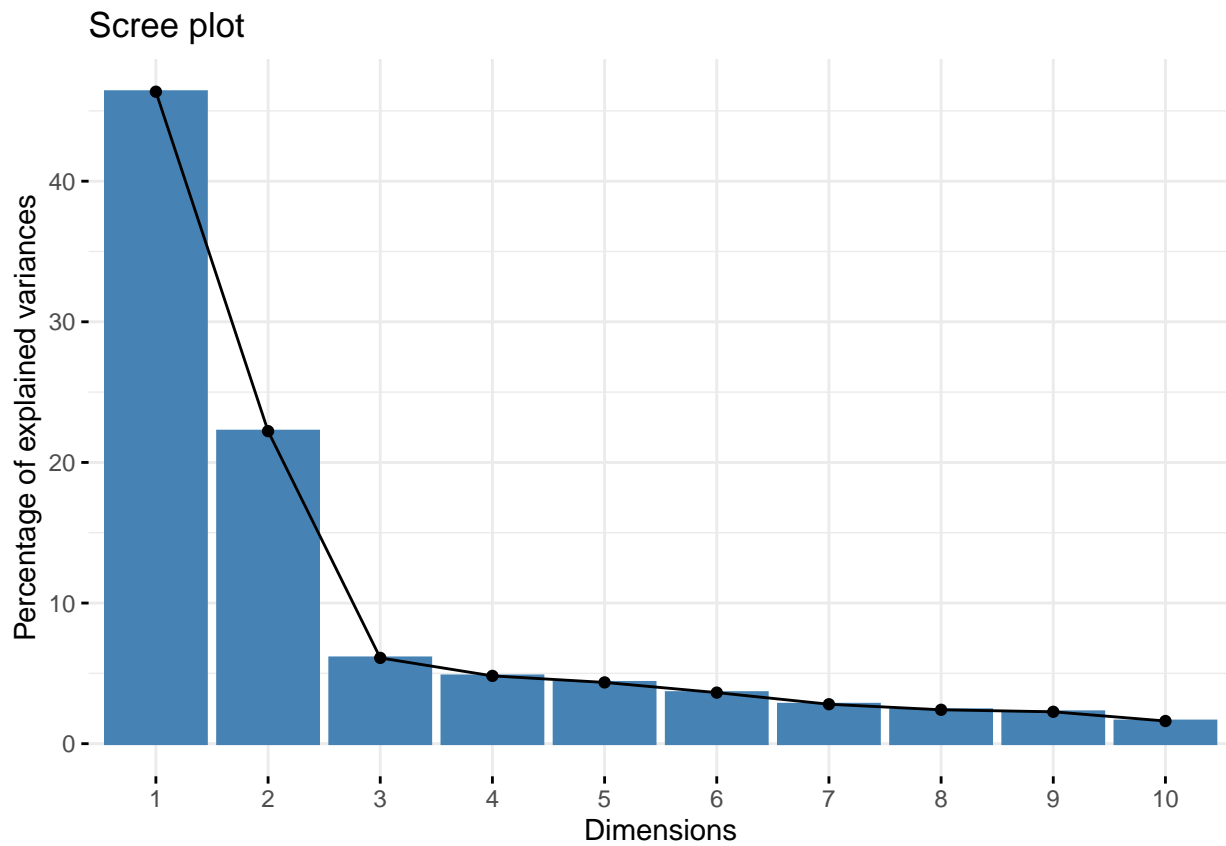
PCA

Conduct PCA on the correlated flood risk variables

```
flood_risk <- fhs_model_df[, 14:35]
```

```
fr_pca <- prcomp(flood_risk[complete.cases(flood_risk),], center = T, scale. = T)
```

```
fr_loadings <- fr_pca$rotation
```

```
fviz_eig(fr_pca)
```

```
summ_pca <- summary(fr_pca)

summ_pca$importance[,1:10]
```

```
##                            PC1      PC2      PC3      PC4       PC5       PC6
## Standard deviation     3.193694 2.211169 1.158406 1.029814 0.9790711 0.8938982
## Proportion of Variance 0.463620 0.222240 0.061000 0.048210 0.0435700 0.0363200
## Cumulative Proportion  0.463620 0.685860 0.746860 0.795060 0.8386300 0.8749500
##                            PC7      PC8      PC9      PC10
## Standard deviation     0.7856323 0.7282944 0.706343 0.5955942
## Proportion of Variance 0.0280600 0.0241100 0.022680 0.0161200
## Cumulative Proportion  0.9030100 0.9271200 0.949800 0.9659200
```

We started out with 22 variables; including seven PC scores would include 90% of the variance.

Printing out the loadings, from most negative to least

```
# First PC Score
fr_loadings[order(fr_loadings[, 1]), 1]
```

```
##      avg_risk_score_all   pct_fs_risk_2020_100   pct_fs_risk_2050_100
##           -0.30983009            -0.29816496            -0.29006089
## avg_risk_score_no_sfha   pct_fs_risk_2020_500   pct_fs_risk_2050_500
##           -0.28928790            -0.27823644            -0.27709006
##      pct_fs_risk_2050_5       pct_floodfactor9     pct_fs_risk_2020_5
##           -0.25924883            -0.23538702            -0.22916964
##      pct_floodfactor10       pct_floodfactor6      pct_floodfactor8
##           -0.21752566            -0.20789454            -0.16744854
##       pct_floodfactor5       pct_floodfactor7   avg_risk_score_sfha
##           -0.16631770            -0.16389343            -0.12728067
##  avg_risk_fsf_2020_500       pct_floodfactor4     avg_risk_score_2_10
##           -0.10519058            -0.09914125            -0.09709154
##  avg_risk_fsf_2020_100       pct_floodfactor3       pct_floodfactor2
##           -0.09035992            -0.08944355            -0.07355485
##       pct_floodfactor1
##            0.27704557
```

The first PC score is very interpretable. Only the loading for pct_floodfactor1 is positive.

```
# Second PC Score
fr_loadings[order(fr_loadings[, 2]), 2]
```

```
##       pct_floodfactor4       pct_floodfactor3   pct_fs_risk_2050_500
##           -0.26345591            -0.26273871            -0.19799053
##       pct_floodfactor2       pct_floodfactor5   pct_fs_risk_2020_500
##           -0.18952970            -0.18411691            -0.17475182
##       pct_floodfactor6   pct_fs_risk_2050_100 avg_risk_score_no_sfha
##           -0.16458545            -0.07060930            -0.07016011
##     avg_risk_score_all   pct_fs_risk_2020_100       pct_floodfactor7
##           -0.03424626             0.00161204             0.08325271
##       pct_floodfactor8       pct_floodfactor9     pct_fs_risk_2050_5
##            0.10940885             0.14048583             0.16973442
##      pct_floodfactor10       pct_floodfactor1     pct_fs_risk_2020_5
##            0.18038281             0.19815384             0.20858359
##    avg_risk_score_sfha  avg_risk_fsf_2020_100  avg_risk_fsf_2020_500
##            0.21875512             0.38240575             0.38780421
##     avg_risk_score_2_10
```

```
##              0.39571585
```

Less interpretable–more of a mix of positive and negative loadings.

```r
# Third PC Score
fr_loadings[order(fr_loadings[, 3]), 3]
```

```
##       pct_floodfactor10       pct_fs_risk_2020_5         pct_floodfactor4
##             -0.40314964             -0.29264410              -0.26919447
##       pct_fs_risk_2050_5 avg_risk_score_no_sfha       avg_risk_score_all
##             -0.22371449             -0.07654446              -0.06789355
##     pct_fs_risk_2020_500     pct_fs_risk_2050_100          pct_floodfactor9
##             -0.05587398             -0.04511104              -0.01296146
##     pct_fs_risk_2050_500        pct_floodfactor1     pct_fs_risk_2020_100
##             -0.01274708              0.01294286               0.02160469
##        pct_floodfactor3      avg_risk_score_2_10       avg_risk_score_sfha
##              0.04316828              0.09131659               0.11472298
## avg_risk_fsf_2020_100    avg_risk_fsf_2020_500          pct_floodfactor2
##              0.11718336              0.11884972               0.13182411
##        pct_floodfactor6        pct_floodfactor5          pct_floodfactor8
##              0.29764585              0.36073167               0.36582151
##        pct_floodfactor7
##              0.43770031
```