Analysis before fitting the CAR model

Alvin Sheng

6/28/2021

```
library(here)
```

```
## here() starts at /Users/Alvin/Documents/NCSU_Spring_2022/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
```

```
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 3.2.1, PROJ 7.2.1
```

```
## 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
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract() masks raster::extract()
```

```
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x dplyr::select()  masks raster::select()
```

```
library(performance)
```

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

Summary Statistics for Table 1 of paper

```
first_var <- 19

summ_stats <- round(t(apply(fhs_model_df[, first_var:ncol(fhs_model_df)], 2, function(vec) {
  c(mean(vec, na.rm = T), sd(vec, na.rm = T), range(vec, na.rm = T))
})),2)

colnames(summ_stats) <- c("mean", "sd", "min", "max")

summ_stats
```

```
##                          mean       sd    min       max
## pct_fs_risk_2020_5       0.03     0.08   0.00      1.00
## pct_fs_risk_2050_5       0.04     0.10   0.00      1.00
## pct_fs_risk_2020_100     0.11     0.15   0.00      1.00
## pct_fs_risk_2050_100     0.12     0.18   0.00      1.00
## pct_fs_risk_2020_500     0.17     0.21   0.00      1.00
## pct_fs_risk_2050_500     0.19     0.22   0.00      1.00
## avg_risk_score_all       1.86     1.14   1.00     10.00
## sd_risk_score_all        1.53     0.77   0.00      6.36
## cv_risk_score_all        0.86     0.32   0.00      1.50
## avg_risk_score_2_10      5.64     1.35   2.00     10.00
## avg_risk_fsf_2020_100    6.76     1.11   3.00     10.00
## avg_risk_fsf_2020_500    5.84     1.32   2.00     10.00
## pct_floodfactor1         0.81     0.22   0.00      1.00
## pct_floodfactor2         0.01     0.04   0.00      1.00
## pct_floodfactor3         0.03     0.06   0.00      1.00
## pct_floodfactor4         0.04     0.09   0.00      1.00
## pct_floodfactor5         0.01     0.03   0.00      1.00
## pct_floodfactor6         0.05     0.08   0.00      1.00
## pct_floodfactor7         0.02     0.03   0.00      1.00
## pct_floodfactor8         0.00     0.01   0.00      1.00
## pct_floodfactor9         0.02     0.05   0.00      1.00
## pct_floodfactor10        0.02     0.06   0.00      1.00
## EP_POV                  15.28    11.93   0.00    100.00
## EP_UNEMP                 6.38     4.67   0.00    100.00
## EP_PCI               32258.07 16848.70  42.00 227064.00
## EP_NOHSDP               13.03    10.56   0.00    100.00
## EP_AGE65                15.98     8.02   0.00    100.00
## EP_AGE17                21.97     6.83   0.00     87.60
## EP_DISABL               13.37     5.88   0.00    100.00
## EP_SNGPNT                9.18     6.44   0.00    100.00
## EP_MINRTY               37.96    30.03   0.00    100.00
## EP_LIMENG                4.13     6.81   0.00    100.00
```

```
## EP_MUNIT                12.25    18.45   0.00    100.00
## EP_MOBILE                6.06    10.76   0.00    100.00
## EP_CROWD                 3.52     5.18   0.00    100.00
## EP_NOVEH                 9.39    12.24   0.00    100.00
## EP_GROUPQ                2.66     9.53   0.00    100.00
## EP_UNINSUR               9.37     7.09   0.00    100.00
## co                       0.36     0.09   0.21      1.93
## no2                     10.20     5.66   1.09     33.08
## o3                      47.32     5.17  29.37     60.51
## pm10                    20.25     5.42   3.88     49.35
## pm25                    10.46     2.32   2.43     18.69
## so2                      2.19     0.98   0.58      9.01
## summer_tmmx            303.09     3.36 289.37    316.04
## winter_tmmx            283.48     7.17 265.42    299.36
## summer_rmax             86.38    11.60  27.90     99.77
## winter_rmax             82.51     7.57  48.82     98.03
## Data_Value_CSMOKING     18.28     5.87   3.20     51.70
## Data_Value_CHD           6.67     2.21   0.50     36.00
## Data_Value_CASTHMA       9.90     1.58   5.40     20.60
## Data_Value_BPHIGH       32.35     7.30   4.90     70.30
## Data_Value_MHLTH        14.26     3.41   5.20     35.50
```

Checking for multicollinearity among the covariates

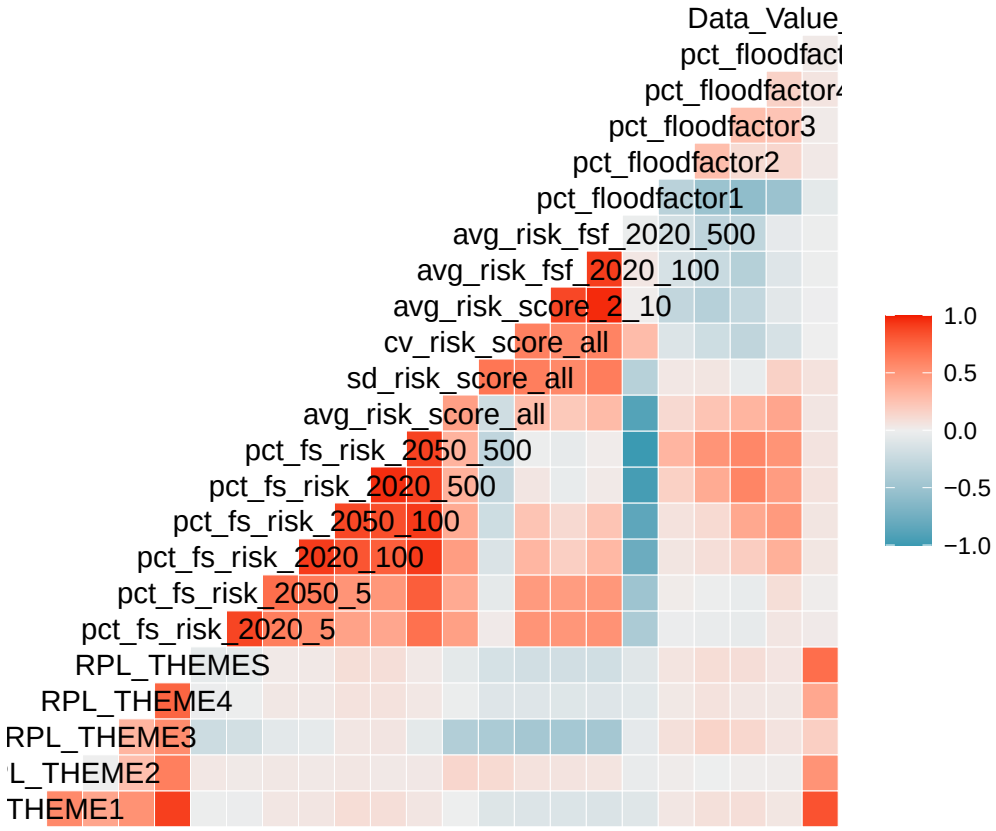
`S.CARleroux()` automatically puts a fixed ridge penalty on the beta coefficients. Therefore, the large number of covariates and multicollinearity would be accounted for.

Actually no, because the penalty is negligible.

Flood risk variables

```
fr_index <- 14:35
```

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




```
flood_cor <- cor(fhs_model_df[complete.cases(fhs_model_df[, c(fr_index, ncol(fhs_model_df))]), c(fr_inde

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

```
##            RPL_THEME1            RPL_THEME2            RPL_THEME3
##           0.837850796           0.520754526           0.173177849
##            RPL_THEME4            RPL_THEMES    pct_fs_risk_2020_5
##           0.417021713           0.711889616           0.028403173
##    pct_fs_risk_2050_5  pct_fs_risk_2020_100  pct_fs_risk_2050_100
##           0.013475915           0.052753614           0.059387849
##  pct_fs_risk_2020_500  pct_fs_risk_2050_500    avg_risk_score_all
##           0.071682035           0.065358829           0.051981560
##     sd_risk_score_all     cv_risk_score_all   avg_risk_score_2_10
##           0.087846659           0.006120672          -0.001212809
## avg_risk_fsf_2020_100 avg_risk_fsf_2020_500      pct_floodfactor1
##          -0.011639662          -0.004086123          -0.065392146
##      pct_floodfactor2      pct_floodfactor3      pct_floodfactor4
##           0.035681138           0.021752121           0.062198689
##      pct_floodfactor5      Data_Value_MHLTH
##           0.031799809           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)
```

```
##    RPL_THEME1          RPL_THEME2          RPL_THEME3          RPL_THEME4
##  Min.   :-0.11718   Min.   :-0.02293   Min.   :-0.452733   Min.   :-0.097837
```

```
##  1st Qu.:-0.00970   1st Qu.: 0.01586   1st Qu.:-0.196935   1st Qu.:-0.006324
##  Median : 0.06513   Median : 0.03147   Median :-0.006871   Median : 0.059932
##  Mean   : 0.15783   Mean   : 0.11934   Mean   :-0.024521   Mean   : 0.114179
##  3rd Qu.: 0.11113   3rd Qu.: 0.08463   3rd Qu.: 0.139624   3rd Qu.: 0.093942
##  Max.   : 0.90379   Max.   : 0.63268   Max.   : 0.541482   Max.   : 0.755098
##  NA's   :1          NA's   :1          NA's   :1           NA's   :1
##    RPL_THEMES       pct_fs_risk_2020_5 pct_fs_risk_2050_5 pct_fs_risk_2020_100
##  Min.   :-0.18408   Min.   :-0.41289   Min.   :-0.50386   Min.   :-0.80510
##  1st Qu.:-0.04362   1st Qu.:-0.02266   1st Qu.:-0.01096   1st Qu.: 0.05552
##  Median : 0.06120   Median : 0.04239   Median : 0.05887   Median : 0.19778
##  Mean   : 0.15438   Mean   : 0.22354   Mean   : 0.23320   Mean   : 0.26975
##  3rd Qu.: 0.11049   3rd Qu.: 0.51637   3rd Qu.: 0.50279   3rd Qu.: 0.57234
##  Max.   : 0.90379   Max.   : 0.88333   Max.   : 0.88333   Max.   : 0.93746
##  NA's   :1          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.   :-0.86713     Min.   :-0.96565     Min.   :-0.99998
##  1st Qu.: 0.05953     1st Qu.: 0.04151     1st Qu.: 0.03342
##  Median : 0.18378     Median : 0.15065     Median : 0.21613
##  Mean   : 0.26879     Mean   : 0.25717     Mean   : 0.25867
##  3rd Qu.: 0.53645     3rd Qu.: 0.51028     3rd Qu.: 0.53219
##  Max.   : 0.93746     Max.   : 0.96558     Max.   : 0.96558
##  NA's   :1            NA's   :1            NA's   :1
##  avg_risk_score_all sd_risk_score_all  cv_risk_score_all  avg_risk_score_2_10
##  Min.   :-0.90136   Min.   :-0.34279   Min.   :-0.45073   Min.   :-0.44130
##  1st Qu.: 0.05493   1st Qu.: 0.03787   1st Qu.:-0.32790   1st Qu.:-0.10486
##  Median : 0.26162   Median : 0.23555   Median :-0.12869   Median : 0.02822
##  Mean   : 0.29314   Mean   : 0.22126   Mean   :-0.04196   Mean   : 0.13763
##  3rd Qu.: 0.62433   3rd Qu.: 0.43395   3rd Qu.: 0.07000   3rd Qu.: 0.43382
##  Max.   : 0.93310   Max.   : 0.60524   Max.   : 0.57834   Max.   : 0.96521
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##  avg_risk_fsf_2020_100 avg_risk_fsf_2020_500 pct_floodfactor1
##  Min.   :-0.45273      Min.   :-0.44132      Min.   :-0.99998
##  1st Qu.:-0.11413      1st Qu.:-0.10173      1st Qu.:-0.56583
##  Median : 0.02152      Median : 0.00823      Median :-0.32949
##  Mean   : 0.12207      Mean   : 0.14724      Mean   :-0.34957
##  3rd Qu.: 0.39986      3rd Qu.: 0.44445      3rd Qu.:-0.06942
##  Max.   : 0.91247      Max.   : 0.96521      Max.   : 0.45065
##  NA's   :1             NA's   :1             NA's   :1
##  pct_floodfactor2   pct_floodfactor3   pct_floodfactor4   pct_floodfactor5
##  Min.   :-0.33348   Min.   :-0.52083   Min.   :-0.57577   Min.   :-0.53600
##  1st Qu.:-0.01161   1st Qu.:-0.02507   1st Qu.:-0.04899   1st Qu.: 0.01589
##  Median : 0.04806   Median : 0.07927   Median : 0.08317   Median : 0.07619
##  Mean   : 0.02925   Mean   : 0.04878   Mean   : 0.05333   Mean   : 0.11105
##  3rd Qu.: 0.12259   3rd Qu.: 0.23453   3rd Qu.: 0.21376   3rd Qu.: 0.25592
##  Max.   : 0.33333   Max.   : 0.52055   Max.   : 0.58559   Max.   : 0.53606
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##  Data_Value_MHLTH
##  Min.   :-0.06539
##  1st Qu.: 0.01554
##  Median : 0.05237
##  Mean   : 0.14395
##  3rd Qu.: 0.08381
##  Max.   : 0.83785
##  NA's   :1
```

Many of the flood risk variables are very correlated.

Using VIF to exclude 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) - 4)]

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



X           <- scale(X) # Scale covariates

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

```
vif(X)
```

```
##                Variables          VIF
## 1             RPL_THEME1    40.461008
## 2             RPL_THEME2    10.360214
## 3             RPL_THEME3    11.494882
## 4             RPL_THEME4    16.357496
## 5             RPL_THEMES   103.082002
## 6     pct_fs_risk_2020_5    13.407814
## 7     pct_fs_risk_2050_5    29.088416
## 8   pct_fs_risk_2020_100    19.391738
## 9   pct_fs_risk_2050_100    21.116009
## 10  pct_fs_risk_2020_500    41.881197
## 11  pct_fs_risk_2050_500 39661.983775
## 12    avg_risk_score_all          Inf
## 13     sd_risk_score_all     6.029168
## 14     cv_risk_score_all     7.470425
## 15   avg_risk_score_2_10    33.682128
## 16 avg_risk_fsf_2020_100     8.144129
## 17 avg_risk_fsf_2020_500    39.746924
## 18      pct_floodfactor2          Inf
## 19      pct_floodfactor3          Inf
## 20      pct_floodfactor4          Inf
## 21      pct_floodfactor5          Inf
## 22      pct_floodfactor6          Inf
## 23      pct_floodfactor7          Inf
## 24      pct_floodfactor8          Inf
## 25      pct_floodfactor9          Inf
## 26     pct_floodfactor10          Inf
## 27                EP_POV     4.497312
## 28              EP_UNEMP     2.588517
## 29                EP_PCI     3.445844
## 30             EP_NOHSDP     6.154116
## 31              EP_AGE65     3.028039
## 32              EP_AGE17     3.822669
## 33             EP_DISABL     4.456609
## 34             EP_SNGPNT     3.967638
```

```
## 35             EP_MINRTY     5.840796
## 36             EP_LIMENG     4.449467
## 37              EP_MUNIT     2.234533
## 38             EP_MOBILE     1.757403
## 39              EP_CROWD     2.974147
## 40              EP_NOVEH     3.278953
## 41             EP_GROUPQ     1.520618
## 42            EP_UNINSUR     2.530316
## 43                    co     9.002284
## 44                   no2    13.724886
## 45                    o3     2.940103
## 46                  pm10     3.778723
## 47                  pm25     5.101244
## 48                   so2     2.645411
## 49           summer_tmmx     4.545334
## 50           winter_tmmx     5.202632
## 51           summer_rmax     3.843705
## 52           winter_rmax     3.279664
## 53   Data_Value_CSMOKING     6.385515
```

```
vifstep(X)
```

```
## 10 variables from the 53 input variables have collinearity problem:
##
## avg_risk_score_all pct_fs_risk_2050_500 RPL_THEMES pct_fs_risk_2020_500 avg_risk_fsf_2020_500 pct_fs
##
## After excluding the collinear variables, the linear correlation coefficients ranges between:
## min correlation ( so2 ~ pct_floodfactor6 ):  -0.0001404145
## max correlation ( pct_floodfactor10 ~ pct_fs_risk_2020_5 ):  0.9069531
##
## ---------- VIFs of the remained variables --------
##                Variables      VIF
## 1             RPL_THEME2 5.899829
## 2             RPL_THEME3 5.502165
## 3             RPL_THEME4 2.238219
## 4     pct_fs_risk_2020_5 8.597118
## 5      sd_risk_score_all 6.221967
## 6      cv_risk_score_all 6.798692
## 7    avg_risk_score_2_10 7.587295
## 8  avg_risk_fsf_2020_100 5.999555
## 9       pct_floodfactor2 1.362569
## 10      pct_floodfactor3 1.561954
## 11      pct_floodfactor4 1.491692
## 12      pct_floodfactor5 1.518450
## 13      pct_floodfactor6 2.096375
## 14      pct_floodfactor7 1.713773
## 15      pct_floodfactor8 1.578252
## 16      pct_floodfactor9 1.941545
## 17     pct_floodfactor10 7.103394
## 18                EP_POV 3.709879
## 19              EP_UNEMP 1.892184
## 20                EP_PCI 2.919854
## 21             EP_NOHSDP 5.686964
## 22              EP_AGE65 2.888338
## 23              EP_AGE17 3.504490
```

```
## 24             EP_DISABL 4.219797
## 25             EP_SNGPNT 4.044651
## 26             EP_MINRTY 5.920815
## 27             EP_LIMENG 4.553059
## 28              EP_MUNIT 2.255493
## 29             EP_MOBILE 1.720212
## 30              EP_CROWD 3.114522
## 31              EP_NOVEH 3.100605
## 32             EP_GROUPQ 1.419496
## 33            EP_UNINSUR 2.520917
## 34                    co 5.103054
## 35                    o3 2.728502
## 36                  pm10 3.734458
## 37                  pm25 4.254676
## 38                   so2 2.558627
## 39           summer_tmmx 4.641896
## 40           winter_tmmx 4.651941
## 41           summer_rmax 3.850532
## 42           winter_rmax 3.051551
## 43   Data_Value_CSMOKING 6.447204
```

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

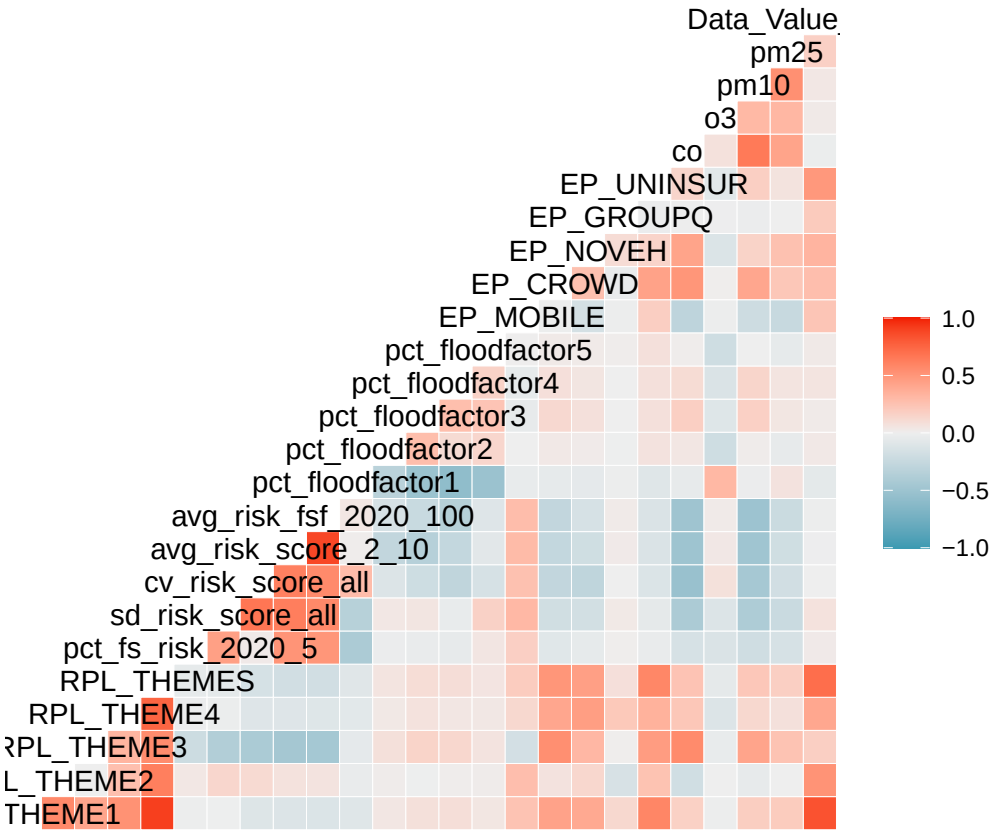
```
collin_var_names <- c("avg_risk_score_all", "pct_fs_risk_2050_500", "pct_fs_risk_2020_500", "avg_risk_fs
```

Correlations among climate related variables: flood risk, pollution, and GRIDMET variables

Excluding variables in collin_var_names

```
climate_var_idx <- c(fr_index, 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))])
```




```
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
```

```
##            RPL_THEME1            RPL_THEME2            RPL_THEME3
##           0.837850796           0.520754526           0.173177849
##            RPL_THEME4            RPL_THEMES    pct_fs_risk_2020_5
##           0.417021713           0.711889616           0.028403173
##     sd_risk_score_all     cv_risk_score_all   avg_risk_score_2_10
##           0.087846659           0.006120672          -0.001212809
## avg_risk_fsf_2020_100      pct_floodfactor1      pct_floodfactor2
##          -0.011639662          -0.065392146           0.035681138
##      pct_floodfactor3      pct_floodfactor4      pct_floodfactor5
##           0.021752121           0.062198689           0.031799809
##             EP_MOBILE             EP_CROWD             EP_NOVEH
##           0.244597997           0.275028726           0.402719725
##             EP_GROUPQ             EP_UNINSUR                    co
##           0.203213252           0.491743362          -0.019936968
##                    o3                  pm10                  pm25
##           0.022837538           0.040438843           0.183587387
##      Data_Value_MHLTH
##           1.000000000
```

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

```
diag(climate_cor) <- NA

summary(climate_cor)
```

```
##    RPL_THEME1         RPL_THEME2           RPL_THEME3         RPL_THEME4
##  Min.   :-0.11718   Min.   :-0.1811692   Min.   :-0.45273   Min.   :-0.11473
##  1st Qu.: 0.02581   1st Qu.:-0.0009937   1st Qu.:-0.09995   1st Qu.: 0.02278
##  Median : 0.14201   Median : 0.0500026   Median : 0.12354   Median : 0.09352
##  Mean   : 0.23007   Mean   : 0.1194813   Mean   : 0.09524   Mean   : 0.17026
##  3rd Qu.: 0.43034   3rd Qu.: 0.2123619   3rd Qu.: 0.33334   3rd Qu.: 0.32170
##  Max.   : 0.90379   Max.   : 0.6326805   Max.   : 0.54645   Max.   : 0.75510
##  NA's   :1          NA's   :1            NA's   :1          NA's   :1
##    RPL_THEMES       pct_fs_risk_2020_5 sd_risk_score_all  cv_risk_score_all
##  Min.   :-0.18408   Min.   :-0.41289   Min.   :-0.34801   Min.   :-0.44973
##  1st Qu.:-0.02895   1st Qu.:-0.09184   1st Qu.:-0.12237   1st Qu.:-0.23247
##  Median : 0.13868   Median :-0.02526   Median : 0.03411   Median :-0.09987
##  Mean   : 0.23580   Mean   : 0.01021   Mean   : 0.04608   Mean   :-0.02627
##  3rd Qu.: 0.51272   3rd Qu.: 0.03083   3rd Qu.: 0.12873   3rd Qu.: 0.09545
##  Max.   : 0.90379   Max.   : 0.52106   Max.   : 0.59397   Max.   : 0.57834
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##  avg_risk_score_2_10 avg_risk_fsf_2020_100 pct_floodfactor1
##  Min.   :-0.45633    Min.   :-0.50374      Min.   :-0.57577
##  1st Qu.:-0.25659    1st Qu.:-0.22761      1st Qu.:-0.16969
##  Median :-0.10286    Median :-0.11107      Median :-0.09726
##  Mean   :-0.01760    Mean   :-0.02159      Mean   :-0.11649
##  3rd Qu.: 0.05343    3rd Qu.: 0.05649      3rd Qu.:-0.02043
##  Max.   : 0.87553    Max.   : 0.87553      Max.   : 0.45065
##  NA's   :1           NA's   :1             NA's   :1
##  pct_floodfactor2    pct_floodfactor3   pct_floodfactor4    pct_floodfactor5
##  Min.   :-0.333478   Min.   :-0.52083   Min.   :-0.575770   Min.   :-0.536001
##  1st Qu.:-0.026455   1st Qu.:-0.03498   1st Qu.:-0.049311   1st Qu.:-0.013618
##  Median : 0.033664   Median : 0.06465   Median : 0.062141   Median : 0.049058
##  Mean   :-0.001767   Mean   : 0.02271   Mean   :-0.003365   Mean   : 0.006787
##  3rd Qu.: 0.058972   3rd Qu.: 0.15573   3rd Qu.: 0.127474   3rd Qu.: 0.072364
##  Max.   : 0.301956   Max.   : 0.30981   Max.   : 0.309812   Max.   : 0.279854
##  NA's   :1           NA's   :1          NA's   :1           NA's   :1
##    EP_MOBILE           EP_CROWD           EP_NOVEH          EP_GROUPQ
##  Min.   :-0.298227   Min.   :-0.28096   Min.   :-0.17304   Min.   :-0.132256
##  1st Qu.:-0.049190   1st Qu.:-0.02907   1st Qu.:-0.06615   1st Qu.:-0.009196
##  Median : 0.006338   Median : 0.10277   Median : 0.10866   Median : 0.006854
##  Mean   : 0.060015   Mean   : 0.13874   Mean   : 0.12214   Mean   : 0.031098
##  3rd Qu.: 0.246570   3rd Qu.: 0.40073   3rd Qu.: 0.26280   3rd Qu.: 0.036784
##  Max.   : 0.300710   Max.   : 0.54119   Max.   : 0.48509   Max.   : 0.231748
##  NA's   :1           NA's   :1          NA's   :1          NA's   :1
##    EP_UNINSUR             co                 o3                 pm10
##  Min.   :-0.12484   Min.   :-0.49052   Min.   :-0.215790   Min.   :-0.50374
##  1st Qu.:-0.03155   1st Qu.:-0.14706   1st Qu.:-0.110054   1st Qu.:-0.08536
##  Median : 0.08995   Median : 0.06708   Median :-0.017517   Median : 0.07602
##  Mean   : 0.15080   Mean   : 0.04958   Mean   :-0.004703   Mean   : 0.05474
##  3rd Qu.: 0.28155   3rd Qu.: 0.20730   3rd Qu.: 0.030316   3rd Qu.: 0.18966
##  Max.   : 0.59252   Max.   : 0.62516   Max.   : 0.348415   Max.   : 0.62516
##  NA's   :1          NA's   :1          NA's   :1           NA's   :1
##       pm25          Data_Value_MHLTH
##  Min.   :-0.23842   Min.   :-0.06539
##  1st Qu.:-0.07074   1st Qu.: 0.02257
##  Median : 0.06051   Median : 0.07502
##  Mean   : 0.06691   Mean   : 0.19585
##  3rd Qu.: 0.19305   3rd Qu.: 0.30695
```

```
##  Max.   : 0.54260   Max.   : 0.83785
##  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) - 4)]

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_fs

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
## -8.6674 -0.4812 -0.0178  0.4524 18.1177
##
## Coefficients:
##                         Estimate Std. Error  t value Pr(>|t|)
## (Intercept)             6.661985   0.003107 2144.130  < 2e-16 ***
## XRPL_THEME1            -0.077443   0.019090   -4.057 4.98e-05 ***
## XRPL_THEME2             0.108505   0.009419   11.520  < 2e-16 ***
## XRPL_THEME3            -0.096476   0.010142   -9.512  < 2e-16 ***
## XRPL_THEME4            -0.008705   0.011953   -0.728 0.466492
## XRPL_THEMES             0.276254   0.030147    9.164  < 2e-16 ***
```

```
## Xpct_fs_risk_2020_5     0.006290   0.008553    0.735 0.462081
## Xsd_risk_score_all      0.049905   0.007696    6.485 8.94e-11 ***
## Xcv_risk_score_all      0.006225   0.007437    0.837 0.402572
## Xavg_risk_score_2_10   -0.033297   0.007510   -4.434 9.28e-06 ***
## Xavg_risk_fsf_2020_100  0.006410   0.006427    0.997 0.318570
## Xpct_floodfactor2      -0.016438   0.003582   -4.590 4.45e-06 ***
## Xpct_floodfactor3      -0.012260   0.003796   -3.230 0.001238 **
## Xpct_floodfactor4      -0.012607   0.003652   -3.452 0.000556 ***
## Xpct_floodfactor5      -0.003501   0.003753   -0.933 0.350834
## Xpct_floodfactor6      -0.005932   0.004252   -1.395 0.162931
## Xpct_floodfactor7      -0.001887   0.003900   -0.484 0.628557
## Xpct_floodfactor8      -0.007726   0.004093   -1.887 0.059109 .
## Xpct_floodfactor9      -0.008132   0.004425   -1.838 0.066121 .
## Xpct_floodfactor10      0.026509   0.007655    3.463 0.000535 ***
## XEP_POV                 0.327431   0.006421   50.996  < 2e-16 ***
## XEP_UNEMP               0.016821   0.005022    3.349 0.000811 ***
## XEP_PCI                 0.013535   0.005756    2.352 0.018699 *
## XEP_NOHSDP              0.201343   0.007759   25.950  < 2e-16 ***
## XEP_AGE65               1.410165   0.005320  265.055  < 2e-16 ***
## XEP_AGE17               0.267978   0.006209   43.161  < 2e-16 ***
## XEP_DISABL              0.243421   0.006188   39.334  < 2e-16 ***
## XEP_SNGPNT             -0.187109   0.005916  -31.627  < 2e-16 ***
## XEP_MINRTY             -0.049112   0.007544   -6.510 7.56e-11 ***
## XEP_LIMENG              0.015637   0.006586    2.374 0.017584 *
## XEP_MUNIT              -0.084596   0.004642  -18.225  < 2e-16 ***
## XEP_MOBILE              0.016936   0.004045    4.187 2.83e-05 ***
## XEP_CROWD              -0.092016   0.005470  -16.821  < 2e-16 ***
## XEP_NOVEH               0.024581   0.005676    4.331 1.49e-05 ***
## XEP_GROUPQ             -0.090131   0.003906  -23.076  < 2e-16 ***
## XEP_UNINSUR             0.132387   0.004850   27.296  < 2e-16 ***
## Xco                     0.022101   0.007131    3.099 0.001940 **
## Xo3                    -0.059906   0.004984  -12.020  < 2e-16 ***
## Xpm10                  -0.015721   0.006198   -2.536 0.011201 *
## Xpm25                  -0.011475   0.006388   -1.796 0.072428 .
## Xso2                    0.086569   0.005153   16.800  < 2e-16 ***
## Xsummer_tmmx            0.126778   0.006574   19.284  < 2e-16 ***
## Xwinter_tmmx            0.065492   0.006560    9.983  < 2e-16 ***
## Xsummer_rmax            0.067309   0.006210   10.839  < 2e-16 ***
## Xwinter_rmax            0.071074   0.005597   12.699  < 2e-16 ***
## XData_Value_CSMOKING    0.827134   0.007613  108.646  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8319 on 71789 degrees of freedom
##   (702 observations deleted due to missingness)
## Multiple R-squared:  0.8581, Adjusted R-squared:  0.858
## F-statistic:  9644 on 45 and 71789 DF,  p-value: < 2.2e-16
```

–>

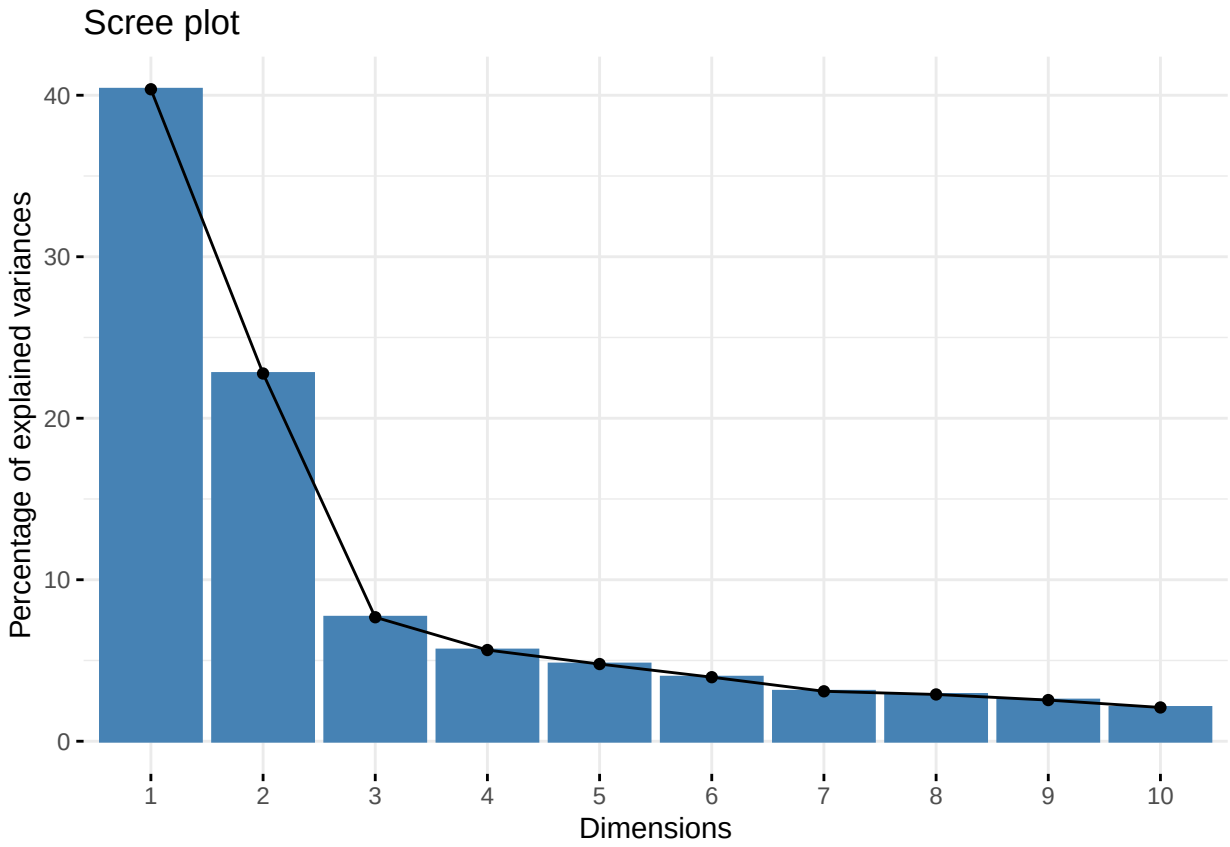
PCA

Conduct PCA on the correlated flood risk variables

```
first_var <- 19

fr_index <- first_var:(first_var + 21)

flood_risk <- fhs_model_df[, fr_index]
```

```
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     2.980165 2.238176 1.299683 1.114667 1.025496 0.9341738
## Proportion of Variance 0.403700 0.227700 0.076780 0.056480 0.047800 0.0396700
## Cumulative Proportion  0.403700 0.631400 0.708180 0.764660 0.812460 0.8521300
##                             PC7       PC8      PC9      PC10
## Standard deviation     0.8249525 0.7983534 0.748906 0.6784936
## Proportion of Variance 0.0309300 0.0289700 0.025490 0.0209300
## Cumulative Proportion  0.8830600 0.9120300 0.937530 0.9584500
```

We started out with 22 variables. Including four PC scores would include 80% of the variance.

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
##           -0.33168172           -0.31448378           -0.31380335
##  pct_fs_risk_2020_500  pct_fs_risk_2050_500    pct_fs_risk_2050_5
##           -0.29916654           -0.29619111           -0.26657190
##    pct_fs_risk_2020_5      pct_floodfactor9     pct_floodfactor10
##           -0.24042633           -0.22217802           -0.21807090
##      pct_floodfactor6     sd_risk_score_all      pct_floodfactor7
##           -0.19907514           -0.17484099           -0.16323272
##      pct_floodfactor8      pct_floodfactor5 avg_risk_fsf_2020_500
##           -0.14190186           -0.14143760           -0.12471527
##   avg_risk_score_2_10 avg_risk_fsf_2020_100      pct_floodfactor4
##           -0.11957319           -0.09942242           -0.09763011
##      pct_floodfactor3      pct_floodfactor2     cv_risk_score_all
##           -0.07886906           -0.04683405            0.08495769
##      pct_floodfactor1
##            0.29616865
```

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
##           -0.26037249           -0.23221303           -0.19136490
##  pct_fs_risk_2020_500      pct_floodfactor6      pct_floodfactor5
##           -0.17012069           -0.16615218           -0.16487957
##      pct_floodfactor2  pct_fs_risk_2050_100    avg_risk_score_all
##           -0.14797990           -0.07046075           -0.03653999
##  pct_fs_risk_2020_100      pct_floodfactor7      pct_floodfactor8
##           -0.01897480            0.05646518            0.08178865
##      pct_floodfactor9    pct_fs_risk_2050_5     pct_floodfactor10
##            0.12622288            0.15849267            0.15921169
##    pct_fs_risk_2020_5      pct_floodfactor1     sd_risk_score_all
##            0.18936889            0.19143421            0.21123041
##     cv_risk_score_all avg_risk_fsf_2020_100 avg_risk_fsf_2020_500
##            0.32161350            0.37692053            0.37889776
##   avg_risk_score_2_10
##            0.38327527
```

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

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

```
##     pct_floodfactor10    pct_fs_risk_2050_5    pct_fs_risk_2020_5
##         -3.974122e-01         -3.277179e-01         -3.127449e-01
##      pct_floodfactor9      pct_floodfactor1    avg_risk_score_all
##         -1.739074e-01         -7.242690e-02         -6.145985e-02
##  pct_fs_risk_2020_100  pct_fs_risk_2050_100      pct_floodfactor4
##          6.653306e-05          1.561162e-02          2.101051e-02
##  pct_fs_risk_2020_500      pct_floodfactor2  pct_fs_risk_2050_500
##          5.586528e-02          6.525560e-02          7.239793e-02
```

```
## avg_risk_fsf_2020_100      pct_floodfactor3   avg_risk_score_2_10
##          9.448015e-02          1.011418e-01          1.100369e-01
## avg_risk_fsf_2020_500      pct_floodfactor8      pct_floodfactor5
##          1.166077e-01          1.516808e-01          2.505114e-01
##      pct_floodfactor6      pct_floodfactor7     cv_risk_score_all
##          2.561423e-01          3.122515e-01          3.797622e-01
##     sd_risk_score_all
##          3.825294e-01
```

```
# Fourth PC Score
fr_loadings[order(fr_loadings[, 4]), 4]
```

```
##      pct_floodfactor8      pct_floodfactor7      pct_floodfactor6
##          -0.281315502          -0.277781005          -0.260406423
##  pct_fs_risk_2020_100  pct_fs_risk_2050_100      pct_floodfactor1
##          -0.179542754          -0.152650065          -0.124261583
##      pct_floodfactor9      pct_floodfactor5    avg_risk_score_all
##          -0.110473643          -0.022257200          -0.017134527
##   avg_risk_score_2_10  pct_fs_risk_2020_500    pct_fs_risk_2050_5
##          -0.007181716           0.006693033           0.036713873
## avg_risk_fsf_2020_500    pct_fs_risk_2020_5  pct_fs_risk_2050_500
##           0.075307417           0.121314451           0.124019886
##      pct_floodfactor4     pct_floodfactor10 avg_risk_fsf_2020_100
##           0.135511378           0.161271275           0.161299133
##     cv_risk_score_all     sd_risk_score_all      pct_floodfactor3
##           0.202762749           0.250355627           0.467353486
##      pct_floodfactor2
##           0.511960031
```

```
# Fifth PC Score
fr_loadings[order(fr_loadings[, 5]), 5]
```

```
##      pct_floodfactor8      pct_floodfactor2      pct_floodfactor7
##           -0.59925282           -0.40980031           -0.36985923
##      pct_floodfactor9      pct_floodfactor3    pct_fs_risk_2050_5
##           -0.22463544           -0.15375374           -0.12115849
##    pct_fs_risk_2020_5      pct_floodfactor1     sd_risk_score_all
##           -0.03398928           -0.02532959            0.01664289
##  pct_fs_risk_2050_500  pct_fs_risk_2020_100    avg_risk_score_all
##            0.02552728            0.03411119            0.03486906
##     cv_risk_score_all avg_risk_fsf_2020_100  pct_fs_risk_2020_500
##            0.05708426            0.08098052            0.09349197
##     pct_floodfactor10  pct_fs_risk_2050_100 avg_risk_fsf_2020_500
##            0.10070678            0.11255958            0.13769357
##   avg_risk_score_2_10      pct_floodfactor5      pct_floodfactor4
##            0.18644632            0.19187611            0.20952684
##      pct_floodfactor6
##            0.25781763
```