Analysis before fitting the CAR model

Alvin Sheng

6/28/2021

```
library(here)
```

```
## Warning in readLines(f, n): line 1 appears to contain an embedded nul

## Warning in readLines(f, n): incomplete final line found on '/Volumes/
## ALVINDRIVE2/flood-risk-health-effects/._flood-risk-health-effects.Rproj'

## here() starts at /Volumes/ALVINDRIVE2/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

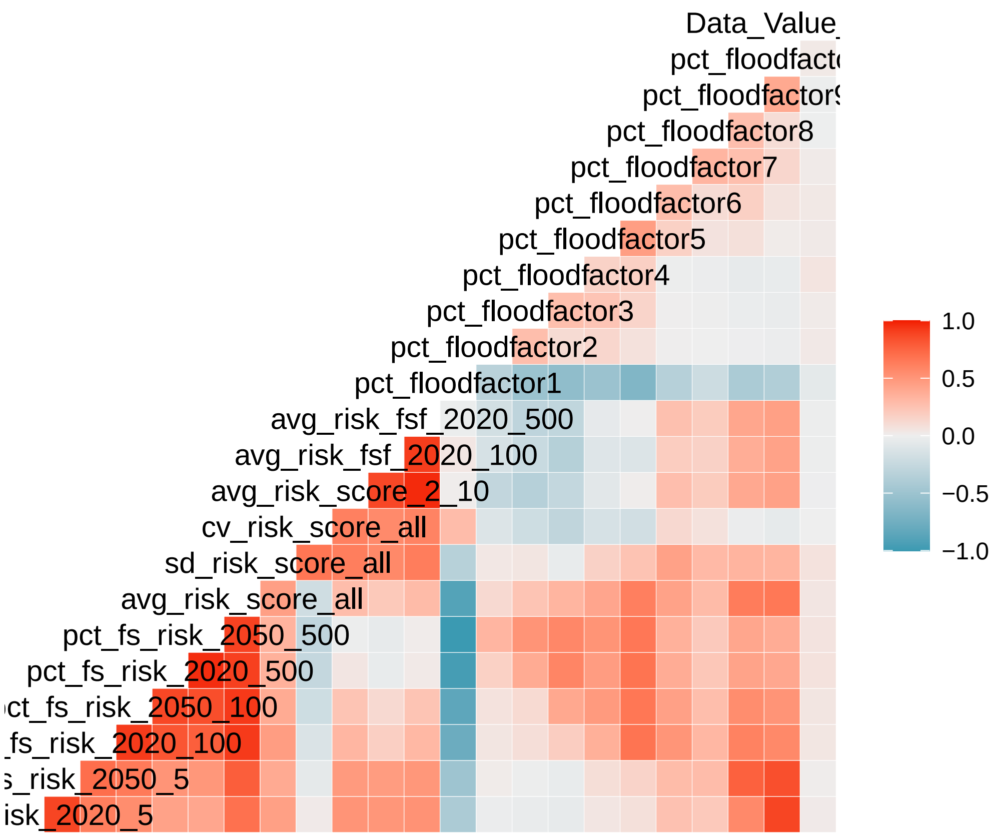
```r
fr_index <- 19:40
```

```r
apply(fhs_model_df[fr_index], 2, function(vec) sd(vec, na.rm = T))
```

```
##    pct_fs_risk_2020_5    pct_fs_risk_2050_5  pct_fs_risk_2020_100
##            0.07588759            0.09713802            0.15288601
##  pct_fs_risk_2050_100  pct_fs_risk_2020_500  pct_fs_risk_2050_500
##            0.17630796            0.20712713            0.22417139
##    avg_risk_score_all     sd_risk_score_all     cv_risk_score_all
##            1.13805698            0.76598522            0.31550045
##   avg_risk_score_2_10 avg_risk_fsf_2020_100 avg_risk_fsf_2020_500
##            1.35167833            1.10638589            1.32059843
##      pct_floodfactor1      pct_floodfactor2      pct_floodfactor3
##            0.22422161            0.03629972            0.06143963
##      pct_floodfactor4      pct_floodfactor5      pct_floodfactor6
##            0.09444138            0.03121215            0.08487768
##      pct_floodfactor7      pct_floodfactor8      pct_floodfactor9
```

```
##            0.02575375           0.01092882           0.04667818
##     pct_floodfactor10
##            0.06209557
```

```
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
```

```
##     pct_fs_risk_2020_5     pct_fs_risk_2050_5   pct_fs_risk_2020_100
##            0.028703467            0.013699636            0.053585772
##   pct_fs_risk_2050_100   pct_fs_risk_2020_500   pct_fs_risk_2050_500
##            0.060164340            0.072378089            0.066051913
##     avg_risk_score_all      sd_risk_score_all      cv_risk_score_all
##            0.052739411            0.088976754            0.006287786
##    avg_risk_score_2_10  avg_risk_fsf_2020_100  avg_risk_fsf_2020_500
##           -0.000661695           -0.011451127           -0.003586298
##       pct_floodfactor1       pct_floodfactor2       pct_floodfactor3
##           -0.066084071            0.035628588            0.022041402
##       pct_floodfactor4       pct_floodfactor5       pct_floodfactor6
##            0.062106455            0.032507082            0.040710421
##       pct_floodfactor7       pct_floodfactor8       pct_floodfactor9
##            0.028293267           -0.004997891           -0.011414654
##      pct_floodfactor10        Data_Value_MHLTH
##            0.033368238            1.000000000
```

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

```r
diag(flood_cor) <- NA

summary(flood_cor)
```

```
##  pct_fs_risk_2020_5 pct_fs_risk_2050_5 pct_fs_risk_2020_100
##  Min.   :-0.41258   Min.   :-0.5036    Min.   :-0.8052
##  1st Qu.: 0.03317   1st Qu.: 0.0418    1st Qu.: 0.1848
##  Median : 0.42470   Median : 0.4285    Median : 0.4121
##  Mean   : 0.32982   Mean   : 0.3539    Mean   : 0.3872
##  3rd Qu.: 0.54557   3rd Qu.: 0.6171    3rd Qu.: 0.6688
##  Max.   : 0.88390   Max.   : 0.8829    Max.   : 0.9373
##  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.   :-0.8672      Min.   :-0.9656      Min.   :-1.0000
##  1st Qu.: 0.1670      1st Qu.: 0.1012      1st Qu.: 0.1173
##  Median : 0.4207      Median : 0.4081      Median : 0.3988
##  Mean   : 0.3762      Mean   : 0.3388      Mean   : 0.3366
##  3rd Qu.: 0.6291      3rd Qu.: 0.5681      3rd Qu.: 0.5650
##  Max.   : 0.9373      Max.   : 0.9656      Max.   : 0.9656
##  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.   :-0.9013    Min.   :-0.3256   Min.   :-0.45146   Min.   :-0.36560
##  1st Qu.: 0.2599    1st Qu.: 0.1610   1st Qu.:-0.31476   1st Qu.:-0.01594
##  Median : 0.4296    Median : 0.3352   Median :-0.11407   Median : 0.25807
##  Mean   : 0.4103    Mean   : 0.3021   Mean   :-0.03583   Mean   : 0.23244
##  3rd Qu.: 0.6805    3rd Qu.: 0.4341   3rd Qu.: 0.04764   3rd Qu.: 0.47516
##  Max.   : 0.9332    Max.   : 0.6054   Max.   : 0.57772   Max.   : 0.96516
##  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.   :-0.35635      Min.   :-0.308819     Min.   :-1.0000
##  1st Qu.:-0.05283      1st Qu.:-0.003222     1st Qu.:-0.6554
##  Median : 0.18534      Median : 0.255205     Median :-0.4204
##  Mean   : 0.20935      Mean   : 0.243471     Mean   :-0.4274
##  3rd Qu.: 0.45731      3rd Qu.: 0.489202     3rd Qu.:-0.2847
##  Max.   : 0.91241      Max.   : 0.965161     Max.   : 0.4514
##  NA's   :1             NA's   :1             NA's   :1
##  pct_floodfactor2   pct_floodfactor3   pct_floodfactor4   pct_floodfactor5
##  Min.   :-0.33290   Min.   :-0.52073   Min.   :-0.57504   Min.   :-0.53499
##  1st Qu.:-0.01982   1st Qu.:-0.03180   1st Qu.:-0.04915   1st Qu.: 0.02091
##  Median : 0.03029   Median : 0.01718   Median :-0.01931   Median : 0.11932
##  Mean   : 0.02048   Mean   : 0.03583   Mean   : 0.03601   Mean   : 0.13565
##  3rd Qu.: 0.12165   3rd Qu.: 0.23844   3rd Qu.: 0.21566   3rd Qu.: 0.34498
##  Max.   : 0.33275   Max.   : 0.52045   Max.   : 0.58485   Max.   : 0.53506
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##  pct_floodfactor6   pct_floodfactor7  pct_floodfactor8   pct_floodfactor9
##  Min.   :-0.68214   Min.   :-0.3593   Min.   :-0.27107   Min.   :-0.42826
##  1st Qu.: 0.04623   1st Qu.: 0.0763   1st Qu.: 0.04401   1st Qu.: 0.01582
##  Median : 0.16419   Median : 0.2623   Median : 0.25821   Median : 0.38830
##  Mean   : 0.19445   Mean   : 0.2218   Mean   : 0.19782   Mean   : 0.29026
##  3rd Qu.: 0.41876   3rd Qu.: 0.3772   3rd Qu.: 0.34150   3rd Qu.: 0.44108
##  Max.   : 0.69827   Max.   : 0.5021   Max.   : 0.46059   Max.   : 0.78917
##  NA's   :1          NA's   :1         NA's   :1          NA's   :1
##  pct_floodfactor10  Data_Value_MHLTH
##  Min.   :-0.38505   Min.   :-0.066084
```

```
##  1st Qu.: 0.02112   1st Qu.: 0.001076
##  Median : 0.36135   Median : 0.030605
##  Mean   : 0.28196   Mean   : 0.027229
##  3rd Qu.: 0.45601   3rd Qu.: 0.053374
##  Max.   : 0.88390   Max.   : 0.088977
##  NA's   :1          NA's   :1
```

Many of the flood risk variables are very correlated.

Using VIF to exclude variables

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

```r
X <- fhs_model_df[, 19:(ncol(fhs_model_df) - 4)]

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



X           <- scale(X) # Scale covariates

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

```r
vif(X)
```

```
##                Variables          VIF
## 1     pct_fs_risk_2020_5     9.356976
## 2     pct_fs_risk_2050_5    29.485793
## 3   pct_fs_risk_2020_100    24.012437
## 4   pct_fs_risk_2050_100    23.335827
## 5   pct_fs_risk_2020_500    44.988462
## 6   pct_fs_risk_2050_500 16973.256253
## 7     avg_risk_score_all          Inf
## 8      sd_risk_score_all     5.958660
## 9      cv_risk_score_all     7.051152
## 10   avg_risk_score_2_10    25.046299
## 11 avg_risk_fsf_2020_100     7.597422
## 12 avg_risk_fsf_2020_500    29.493051
## 13       pct_floodfactor2          Inf
## 14       pct_floodfactor3          Inf
## 15       pct_floodfactor4          Inf
## 16       pct_floodfactor5          Inf
## 17       pct_floodfactor6          Inf
## 18       pct_floodfactor7          Inf
## 19       pct_floodfactor8          Inf
## 20       pct_floodfactor9          Inf
## 21      pct_floodfactor10          Inf
## 22                EP_POV     3.739135
## 23              EP_UNEMP     1.951955
## 24                EP_PCI     2.927168
## 25             EP_NOHSDP     5.475764
## 26              EP_AGE65     2.303764
## 27              EP_AGE17     2.826034
```

```
## 28              EP_DISABL     2.759972
## 29              EP_SNGPNT     2.522560
## 30              EP_MINRTY     4.034629
## 31              EP_LIMENG     3.777177
## 32               EP_MUNIT     2.001206
## 33              EP_MOBILE     1.634880
## 34               EP_CROWD     2.452599
## 35               EP_NOVEH     3.018122
## 36              EP_GROUPQ     1.388930
## 37             EP_UNINSUR     2.332799
## 38                     co     9.469019
## 39                    no2    14.269833
## 40                     o3     2.704467
## 41                   pm10     3.685572
## 42                   pm25     5.107574
## 43                    so2     2.563143
## 44            summer_tmmx     4.318664
## 45            winter_tmmx     5.156564
## 46            summer_rmax     3.801672
## 47            winter_rmax     3.155196
## 48   Data_Value_CSMOKING     6.207479
```

```
vifstep(X)
```

```
## 8 variables from the 48 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 ( pct_floodfactor9 ~ pct_floodfactor2 ):  0.0002746726
## max correlation ( avg_risk_fsf_2020_100 ~ avg_risk_score_2_10 ):  0.8822811
##
## ---------- VIFs of the remained variables --------
##                Variables      VIF
## 1      pct_fs_risk_2020_5 5.238394
## 2       sd_risk_score_all 5.431663
## 3       cv_risk_score_all 7.456191
## 4     avg_risk_score_2_10 8.008611
## 5   avg_risk_fsf_2020_100 6.309733
## 6       pct_floodfactor2 1.368531
## 7       pct_floodfactor3 1.603086
## 8       pct_floodfactor4 1.555496
## 9       pct_floodfactor5 1.458488
## 10      pct_floodfactor6 2.036177
## 11      pct_floodfactor7 1.630830
## 12      pct_floodfactor8 1.495920
## 13      pct_floodfactor9 2.079738
## 14     pct_floodfactor10 4.498230
## 15                EP_POV 3.595223
## 16              EP_UNEMP 2.015822
## 17                EP_PCI 2.984099
## 18             EP_NOHSDP 5.633798
## 19              EP_AGE65 2.384307
## 20              EP_AGE17 2.790759
## 21             EP_DISABL 2.752264
```

```
## 22             EP_SNGPNT 2.662548
## 23             EP_MINRTY 3.567756
## 24             EP_LIMENG 3.922705
## 25              EP_MUNIT 1.923146
## 26             EP_MOBILE 1.649403
## 27              EP_CROWD 2.760508
## 28              EP_NOVEH 2.714328
## 29             EP_GROUPQ 1.466398
## 30            EP_UNINSUR 2.395901
## 31                    co 4.351053
## 32                    o3 2.721062
## 33                  pm10 3.485918
## 34                  pm25 4.243256
## 35                   so2 2.484211
## 36           summer_tmmx 4.393448
## 37           winter_tmmx 4.575223
## 38           summer_rmax 3.689015
## 39           winter_rmax 2.871968
## 40   Data_Value_CSMOKING 6.144305
```

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

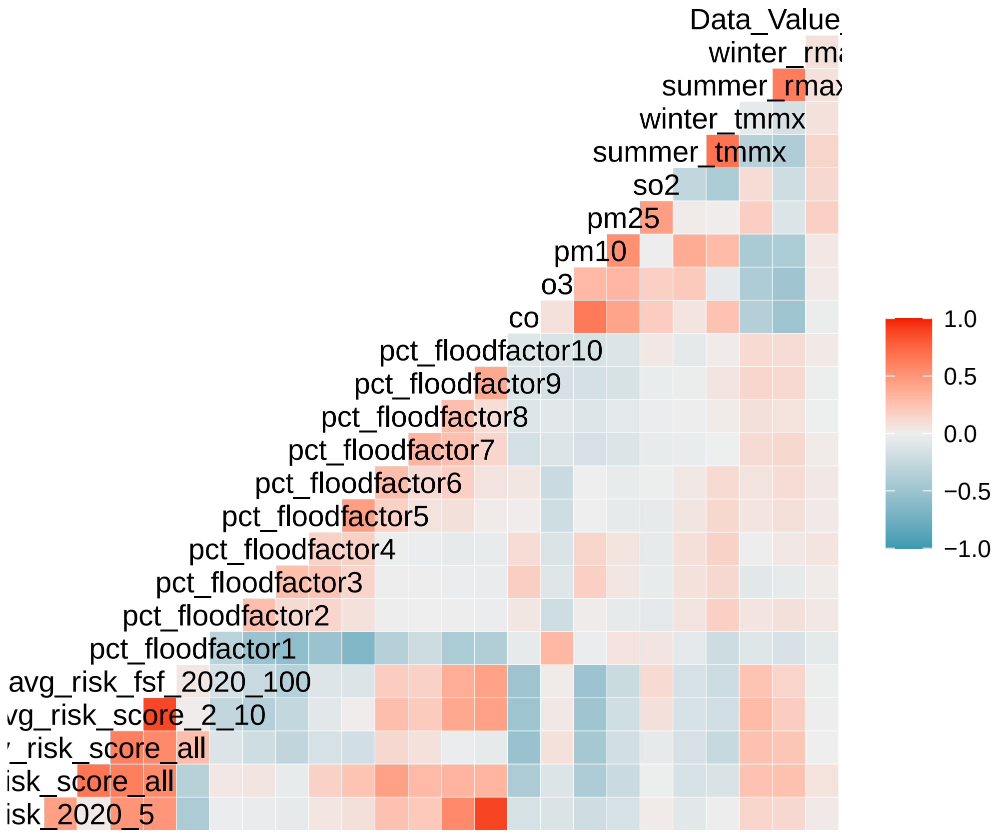
```
collin_var_names <- c("avg_risk_score_all", "pct_fs_risk_2050_500", "pct_fs_risk_2020_500", "avg_risk_f
```

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

Excluding variables in collin_var_names

```
climate_var_idx <- c(fr_index, 57:66)

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))])
```




```
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     sd_risk_score_all     cv_risk_score_all
##          0.0299804997          0.0889601742          0.0060502492
##    avg_risk_score_2_10 avg_risk_fsf_2020_100       pct_floodfactor1
##         -0.0004361047         -0.0112722648         -0.0665488462
##      pct_floodfactor2      pct_floodfactor3      pct_floodfactor4
##          0.0356060183          0.0220103942          0.0620835478
##      pct_floodfactor5      pct_floodfactor6      pct_floodfactor7
##          0.0324997455          0.0407283782          0.0287804262
##      pct_floodfactor8      pct_floodfactor9     pct_floodfactor10
##         -0.0048493860         -0.0112121004          0.0349543674
##                    co                    o3                  pm10
##         -0.0196370204          0.0226626739          0.0405151007
##                  pm25                   so2           summer_tmmx
##          0.1832766126          0.1537557332          0.1454940393
##           winter_tmmx           summer_rmax           winter_rmax
##          0.0731103518          0.0944951518          0.0855385378
##      Data_Value_MHLTH
##          1.0000000000
```

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

```
diag(climate_cor) <- NA

summary(climate_cor)
```

```
##  pct_fs_risk_2020_5 sd_risk_score_all cv_risk_score_all  avg_risk_score_2_10
##  Min.   :-0.40902   Min.   :-0.3474   Min.   :-0.44990   Min.   :-0.45656
##  1st Qu.:-0.05581   1st Qu.:-0.1228   1st Qu.:-0.26268   1st Qu.:-0.17685
##  Median : 0.03657   Median : 0.1166   Median :-0.04901   Median : 0.04197
##  Mean   : 0.11607   Mean   : 0.1228   Mean   :-0.01306   Mean   : 0.08615
##  3rd Qu.: 0.27656   3rd Qu.: 0.3371   3rd Qu.: 0.11207   3rd Qu.: 0.29873
##  Max.   : 0.88042   Max.   : 0.5955   Max.   : 0.58195   Max.   : 0.87512
##  NA's   :1          NA's   :1         NA's   :1          NA's   :1
##  avg_risk_fsf_2020_100 pct_floodfactor1   pct_floodfactor2
##  Min.   :-0.50351      Min.   :-0.68385   Min.   :-0.333951
##  1st Qu.:-0.16835      1st Qu.:-0.38829   1st Qu.:-0.051353
##  Median : 0.04115      Median :-0.17383   Median : 0.013113
##  Mean   : 0.07535      Mean   :-0.19129   Mean   :-0.001026
##  3rd Qu.: 0.28041      3rd Qu.:-0.03059   3rd Qu.: 0.078299
##  Max.   : 0.87512      Max.   : 0.44949   Max.   : 0.301375
##  NA's   :1             NA's   :1          NA's   :1
##  pct_floodfactor3    pct_floodfactor4   pct_floodfactor5   pct_floodfactor6
##  Min.   :-0.522389   Min.   :-0.57674   Min.   :-0.53652   Min.   :-0.6838548
##  1st Qu.:-0.052386   1st Qu.:-0.04865   1st Qu.:-0.04306   1st Qu.:-0.0004318
##  Median : 0.003107   Median :-0.01916   Median : 0.04986   Median : 0.0689923
##  Mean   : 0.002413   Mean   :-0.02058   Mean   : 0.03239   Mean   : 0.0400112
##  3rd Qu.: 0.150393   3rd Qu.: 0.13801   3rd Qu.: 0.14366   3rd Qu.: 0.1535915
##  Max.   : 0.310142   Max.   : 0.31014   Max.   : 0.46674   Max.   : 0.4667402
##  NA's   :1           NA's   :1          NA's   :1          NA's   :1
##  pct_floodfactor7   pct_floodfactor8   pct_floodfactor9   pct_floodfactor10
##  Min.   :-0.35804   Min.   :-0.27042   Min.   :-0.42623   Min.   :-0.38138
##  1st Qu.:-0.02109   1st Qu.:-0.02231   1st Qu.:-0.05628   1st Qu.:-0.06516
##  Median : 0.04330   Median : 0.03359   Median : 0.03625   Median : 0.02956
##  Mean   : 0.07881   Mean   : 0.07684   Mean   : 0.09221   Mean   : 0.08918
##  3rd Qu.: 0.21338   3rd Qu.: 0.16632   3rd Qu.: 0.28783   3rd Qu.: 0.14090
##  Max.   : 0.45756   Max.   : 0.45756   Max.   : 0.58777   Max.   : 0.88042
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##        co                 o3                pm10               pm25
##  Min.   :-0.49037   Min.   :-0.50981   Min.   :-0.50351   Min.   :-0.22305
##  1st Qu.:-0.18352   1st Qu.:-0.13751   1st Qu.:-0.23089   1st Qu.:-0.13091
##  Median :-0.05265   Median :-0.09970   Median :-0.02814   Median :-0.03210
##  Mean   :-0.04405   Mean   :-0.04017   Mean   :-0.02623   Mean   : 0.03357
##  3rd Qu.: 0.11930   3rd Qu.: 0.08466   3rd Qu.: 0.17628   3rd Qu.: 0.09245
##  Max.   : 0.62502   Max.   : 0.34745   Max.   : 0.62502   Max.   : 0.54279
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##       so2            summer_tmmx         winter_tmmx        summer_rmax
##  Min.   :-0.41123   Min.   :-0.354306   Min.   :-0.41123   Min.   :-0.39896
##  1st Qu.:-0.04441   1st Qu.:-0.148715   1st Qu.:-0.13020   1st Qu.:-0.03477
##  Median : 0.02180   Median :-0.010138   Median : 0.01709   Median : 0.08962
##  Mean   : 0.01892   Mean   : 0.001591   Mean   : 0.02280   Mean   : 0.04483
##  3rd Qu.: 0.11396   3rd Qu.: 0.072685   3rd Qu.: 0.14145   3rd Qu.: 0.14923
##  Max.   : 0.46857   Max.   : 0.692370   Max.   : 0.69237   Max.   : 0.59653
##  NA's   :1          NA's   :1           NA's   :1          NA's   :1
##   winter_rmax        Data_Value_MHLTH
##  Min.   :-0.509809   Min.   :-0.066549
##  1st Qu.:-0.132134   1st Qu.: 0.004429
##  Median : 0.079918   Median : 0.033727
##  Mean   :-0.007747   Mean   : 0.044439
##  3rd Qu.: 0.111153   3rd Qu.: 0.076217
```

```
## Max.   : 0.596532   Max.   : 0.183277
## NA's   :1           NA's   :1
```

Climate variables other than flood risk are not too correlated.

Non-spatial modeling

```r
Y <- fhs_model_df$Data_Value_CHD

X <- fhs_model_df[, 19:(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_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)
```

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

```r
summary(fhs_lm)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.8599 -0.4803 -0.0189  0.4575 17.7353
##
## Coefficients:
##                           Estimate Std. Error  t value Pr(>|t|)
## (Intercept)              6.6601517  0.0031357 2123.978  < 2e-16 ***
## Xpct_fs_risk_2020_5      0.0041458  0.0086312    0.480 0.630996
## Xsd_risk_score_all       0.0552776  0.0077439    7.138 9.54e-13 ***
## Xcv_risk_score_all       0.0141664  0.0074888    1.892 0.058538 .
## Xavg_risk_score_2_10    -0.0320944  0.0075740   -4.237 2.26e-05 ***
## Xavg_risk_fsf_2020_100   0.0064095  0.0064859    0.988 0.323050
```

```
## Xpct_floodfactor2      -0.0175671  0.0036160   -4.858 1.19e-06 ***
## Xpct_floodfactor3      -0.0116605  0.0038324   -3.043 0.002346 **
## Xpct_floodfactor4      -0.0112754  0.0036872   -3.058 0.002229 **
## Xpct_floodfactor5      -0.0025495  0.0037888   -0.673 0.501012
## Xpct_floodfactor6      -0.0077272  0.0042902   -1.801 0.071688 .
## Xpct_floodfactor7      -0.0002381  0.0039344   -0.061 0.951752
## Xpct_floodfactor8      -0.0081307  0.0041329   -1.967 0.049152 *
## Xpct_floodfactor9      -0.0099330  0.0044657   -2.224 0.026133 *
## Xpct_floodfactor10      0.0252324  0.0077199    3.268 0.001082 **
## XEP_POV                 0.3390289  0.0059369   57.105  < 2e-16 ***
## XEP_UNEMP               0.0157048  0.0043573    3.604 0.000313 ***
## XEP_PCI                -0.0258376  0.0052305   -4.940 7.84e-07 ***
## XEP_NOHSDP              0.2126295  0.0074081   28.702  < 2e-16 ***
## XEP_AGE65               1.4712867  0.0047719  308.324  < 2e-16 ***
## XEP_AGE17               0.3361792  0.0053876   62.398  < 2e-16 ***
## XEP_DISABL              0.3453552  0.0051798   66.673  < 2e-16 ***
## XEP_SNGPNT             -0.1013951  0.0050818  -19.953  < 2e-16 ***
## XEP_MINRTY             -0.0666869  0.0059761  -11.159  < 2e-16 ***
## XEP_LIMENG             -0.0073262  0.0062244   -1.177 0.239195
## XEP_MUNIT              -0.0592066  0.0044919  -13.181  < 2e-16 ***
## XEP_MOBILE              0.0416902  0.0039966   10.431  < 2e-16 ***
## XEP_CROWD              -0.0675969  0.0053378  -12.664  < 2e-16 ***
## XEP_NOVEH               0.0455598  0.0056279    8.095 5.80e-16 ***
## XEP_GROUPQ             -0.0756168  0.0038800  -19.489  < 2e-16 ***
## XEP_UNINSUR             0.1515184  0.0048360   31.331  < 2e-16 ***
## Xco                     0.0175963  0.0071454    2.463 0.013796 *
## Xo3                    -0.0628817  0.0050313  -12.498  < 2e-16 ***
## Xpm10                  -0.0072726  0.0062277   -1.168 0.242902
## Xpm25                  -0.0073628  0.0064447   -1.142 0.253268
## Xso2                    0.0821766  0.0051917   15.829  < 2e-16 ***
## Xsummer_tmmx            0.1184252  0.0066269   17.870  < 2e-16 ***
## Xwinter_tmmx            0.0633490  0.0066074    9.588  < 2e-16 ***
## Xsummer_rmax            0.0599815  0.0062598    9.582  < 2e-16 ***
## Xwinter_rmax            0.0821721  0.0056402   14.569  < 2e-16 ***
## XData_Value_CSMOKING    0.8362653  0.0075716  110.447  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.84 on 71794 degrees of freedom
##   (702 observations deleted due to missingness)
## Multiple R-squared:  0.8553, Adjusted R-squared:  0.8552
## F-statistic: 1.061e+04 on 40 and 71794 DF,  p-value: < 2.2e-16
```

–>

PCA with Centering but no Scaling beforehand

Do PCA *without* scaling beforehand, and use biplots, etc. to compare results with those in the last section. I think just scaling all covariates once, *after* PCA, will lead to more interpretable results

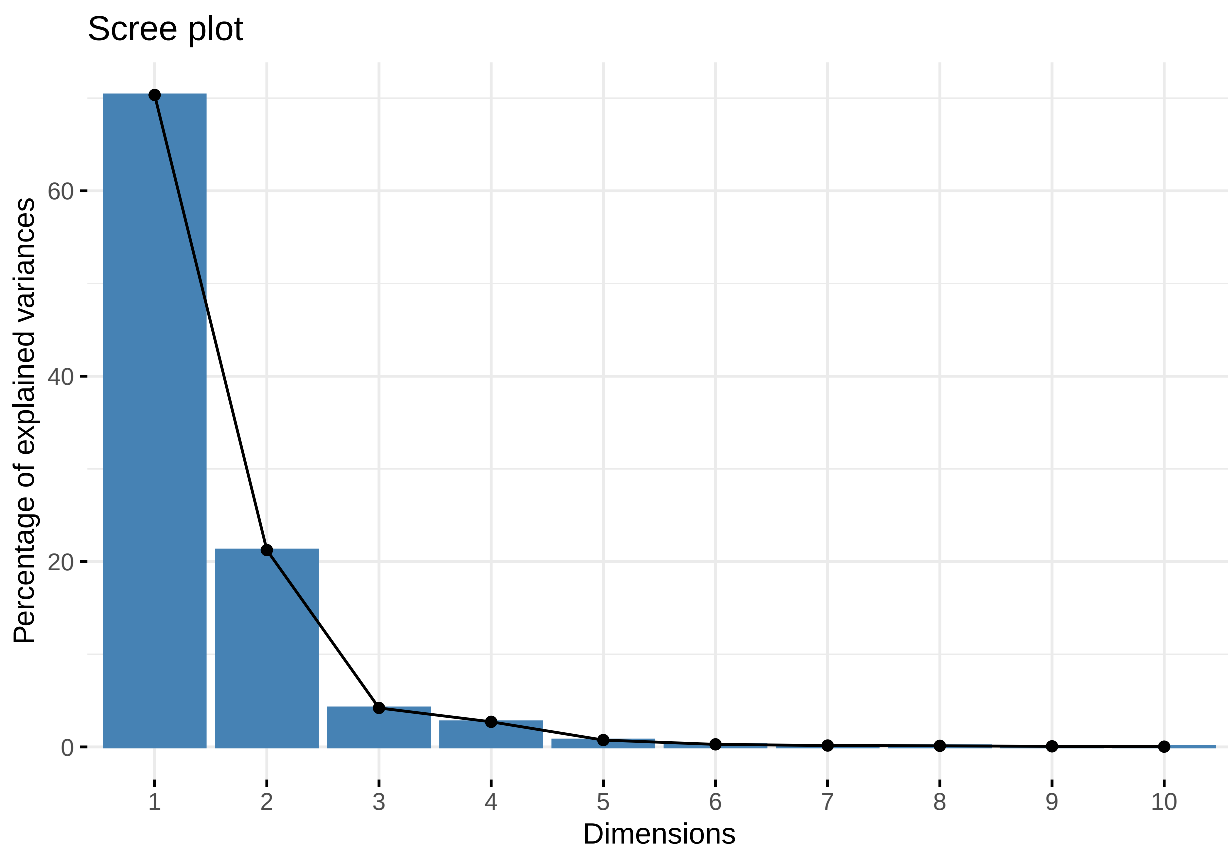
Conduct PCA on the correlated flood risk variables

```
first_var <- 19

fr_index <- first_var:(first_var + 21)
```

```r
flood_risk <- fhs_model_df[, fr_index]
```

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

fr_loadings <- fr_pca$rotation
```

```r
fviz_eig(fr_pca)
```




```r
summ_pca <- summary(fr_pca)

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

```
##                             PC1      PC2       PC3       PC4       PC5
## Standard deviation     2.156094 1.184867 0.5271685 0.4232675 0.2213399
## Proportion of Variance 0.703430 0.212440 0.0420500 0.0271100 0.0074100
## Cumulative Proportion  0.703430 0.915870 0.9579200 0.9850300 0.9924500
##                              PC6       PC7        PC8       PC9       PC10
## Standard deviation     0.1354879 0.1002485 0.08885766 0.0664214 0.04515333
## Proportion of Variance 0.0027800 0.0015200 0.00119000 0.0006700 0.00031000
## Cumulative Proportion  0.9952200 0.9967400 0.99794000 0.9986100 0.99891000
```

We started out with 22 variables. Including two PC scores would include >90% of the variance. Perhaps I can also look at the top 5 PCs, to get > 99% variance explained.

Printing out the loadings, from most negative to least

```r
# First PC Score
fr_loadings[, 1]
```

```
##    pct_fs_risk_2020_5    pct_fs_risk_2050_5  pct_fs_risk_2020_100
##          -0.021240339          -0.026200603          -0.027287959
##  pct_fs_risk_2050_100  pct_fs_risk_2020_500  pct_fs_risk_2050_500
##          -0.025890403          -0.012937437          -0.011239303
##    avg_risk_score_all     sd_risk_score_all     cv_risk_score_all
##          -0.206655116          -0.224161348          -0.065925907
##   avg_risk_score_2_10 avg_risk_fsf_2020_100 avg_risk_fsf_2020_500
##          -0.581194182          -0.477386598          -0.577512216
##      pct_floodfactor1      pct_floodfactor2      pct_floodfactor3
##           0.011215117           0.002775862           0.006705547
##      pct_floodfactor4      pct_floodfactor5      pct_floodfactor6
##           0.010804712           0.000436761          -0.001712055
##      pct_floodfactor7      pct_floodfactor8      pct_floodfactor9
##          -0.003673378          -0.001119415          -0.010331553
##     pct_floodfactor10
##          -0.015101598
```

The first PC score is very interpretable. Only the loadings for the first five pct_floodfactor variables are positive.

```
# Second PC Score
fr_loadings[, 2]
```

```
##    pct_fs_risk_2020_5    pct_fs_risk_2050_5  pct_fs_risk_2020_100
##           0.030877735           0.048298401           0.110524027
##  pct_fs_risk_2050_100  pct_fs_risk_2020_500  pct_fs_risk_2050_500
##           0.130411155           0.162847406           0.176634178
##    avg_risk_score_all     sd_risk_score_all     cv_risk_score_all
##           0.882803822           0.144210669          -0.127097694
##   avg_risk_score_2_10 avg_risk_fsf_2020_100 avg_risk_fsf_2020_500
##          -0.130010196          -0.161871141          -0.117887932
##      pct_floodfactor1      pct_floodfactor2      pct_floodfactor3
##          -0.176663004           0.006636052           0.019301215
##      pct_floodfactor4      pct_floodfactor5      pct_floodfactor6
##           0.035596848           0.012790780           0.048754950
##      pct_floodfactor7      pct_floodfactor8      pct_floodfactor9
##           0.007900621           0.001889291           0.018931063
##     pct_floodfactor10
##           0.024862184
```

The second PC score only has negative loadings for pct_floodfactor1 and some of the avg_risk_score variables.

```
round(fr_loadings[, 1:2], digits = 2)
```

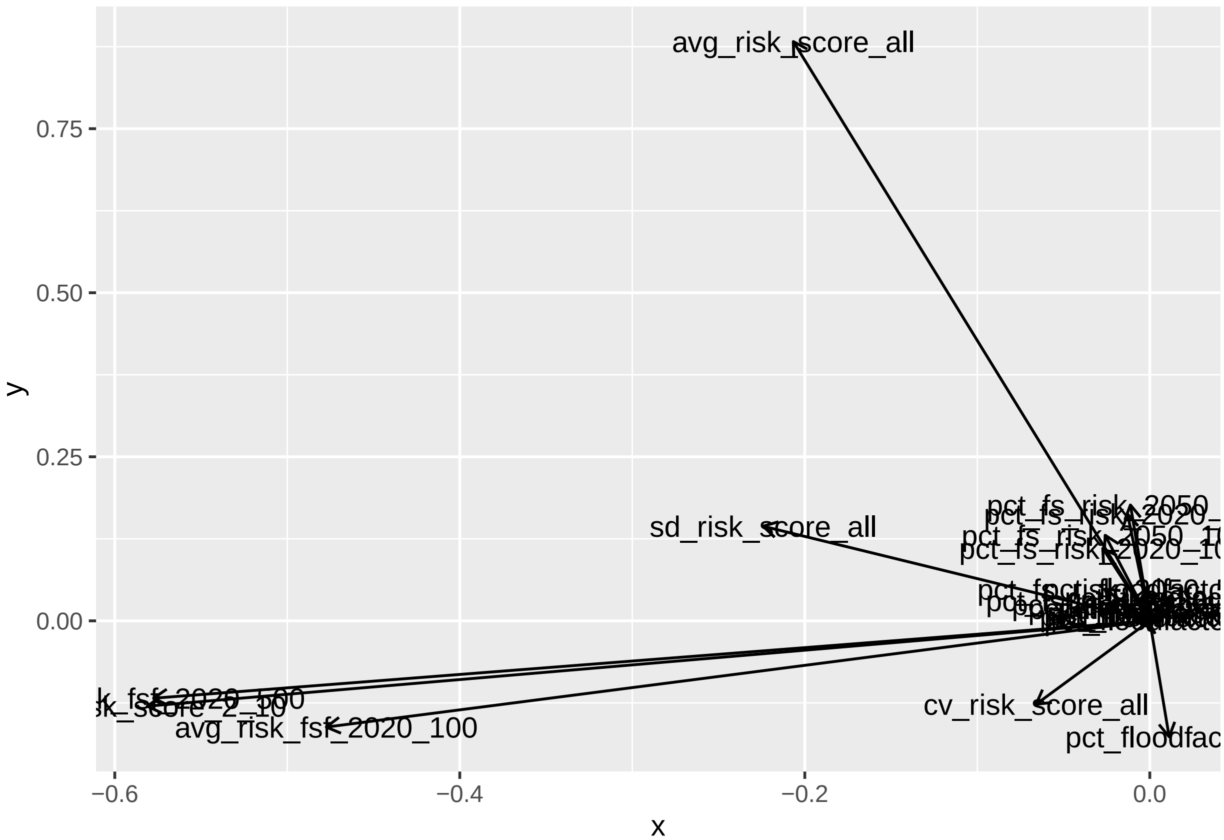
```
##                         PC1   PC2
## pct_fs_risk_2020_5    -0.02  0.03
## pct_fs_risk_2050_5    -0.03  0.05
## pct_fs_risk_2020_100  -0.03  0.11
## pct_fs_risk_2050_100  -0.03  0.13
## pct_fs_risk_2020_500  -0.01  0.16
## pct_fs_risk_2050_500  -0.01  0.18
## avg_risk_score_all    -0.21  0.88
## sd_risk_score_all     -0.22  0.14
## cv_risk_score_all     -0.07 -0.13
## avg_risk_score_2_10   -0.58 -0.13
```

```
## avg_risk_fsf_2020_100 -0.48 -0.16
## avg_risk_fsf_2020_500 -0.58 -0.12
## pct_floodfactor1       0.01 -0.18
## pct_floodfactor2       0.00  0.01
## pct_floodfactor3       0.01  0.02
## pct_floodfactor4       0.01  0.04
## pct_floodfactor5       0.00  0.01
## pct_floodfactor6       0.00  0.05
## pct_floodfactor7       0.00  0.01
## pct_floodfactor8       0.00  0.00
## pct_floodfactor9      -0.01  0.02
## pct_floodfactor10     -0.02  0.02
```

```r
round(fr_loadings[, 1:8], digits = 2)
```

```
##                         PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8
## pct_fs_risk_2020_5    -0.02  0.03  0.02 -0.02 -0.02  0.23 -0.04 -0.24
## pct_fs_risk_2050_5    -0.03  0.05  0.04 -0.02 -0.02  0.30  0.02 -0.25
## pct_fs_risk_2020_100  -0.03  0.11  0.02  0.04 -0.02  0.07  0.11  0.40
## pct_fs_risk_2050_100  -0.03  0.13  0.03  0.04  0.01 -0.10  0.07  0.34
## pct_fs_risk_2020_500  -0.01  0.16 -0.01  0.02 -0.06 -0.39 -0.01 -0.03
## pct_fs_risk_2050_500  -0.01  0.18 -0.02 -0.03  0.08 -0.44 -0.04 -0.11
## avg_risk_score_all    -0.21  0.88  0.16 -0.06 -0.05  0.20  0.16 -0.06
## sd_risk_score_all     -0.22  0.14 -0.93  0.07  0.01  0.10 -0.23  0.03
## cv_risk_score_all     -0.07 -0.13 -0.25  0.06 -0.04 -0.09  0.93 -0.17
## avg_risk_score_2_10   -0.58 -0.13  0.17  0.52 -0.56 -0.08 -0.08 -0.03
## avg_risk_fsf_2020_100 -0.48 -0.16  0.02 -0.83 -0.23 -0.06  0.00  0.05
## avg_risk_fsf_2020_500 -0.58 -0.12  0.14  0.14  0.78  0.03  0.00  0.00
## pct_floodfactor1       0.01 -0.18  0.02  0.03 -0.08  0.44  0.04  0.11
## pct_floodfactor2       0.00  0.01 -0.01 -0.02  0.05 -0.03  0.00 -0.04
## pct_floodfactor3       0.01  0.02 -0.02 -0.03  0.04 -0.13 -0.05 -0.17
## pct_floodfactor4       0.01  0.04  0.00  0.02  0.01 -0.37 -0.10 -0.36
## pct_floodfactor5       0.00  0.01  0.00  0.00  0.01 -0.05  0.00  0.08
## pct_floodfactor6       0.00  0.05  0.00  0.02  0.01 -0.15  0.07  0.57
## pct_floodfactor7       0.00  0.01 -0.01  0.00  0.00  0.01  0.01  0.06
## pct_floodfactor8       0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.01
## pct_floodfactor9      -0.01  0.02  0.01 -0.01 -0.01  0.09  0.04 -0.03
## pct_floodfactor10     -0.02  0.02  0.02 -0.01 -0.02  0.20 -0.02 -0.21
```

```r
# Extract loadings of the variables
fr_loadings_df <- data.frame(Variables = rownames(fr_pca$rotation), fr_pca$rotation)

    # Plot
ggplot(fr_loadings_df) +
  geom_segment(data = fr_loadings_df, aes(x = 0, y = 0, xend = PC1,
     yend = PC2), arrow = arrow(length = unit(1/2, "picas")),
     color = "black") +
  annotate("text", x = (fr_loadings_df$PC1), y = (fr_loadings_df$PC2),
     label = fr_loadings_df$Variables)
```



Re-checking the Model Diagnostics with Flood Risk PCs

Noticed that NO2 has a VIF greater than 10. This variable should probably be removed from the analysis. Let's see what effect this has on the coefficients of the other pollution variables.

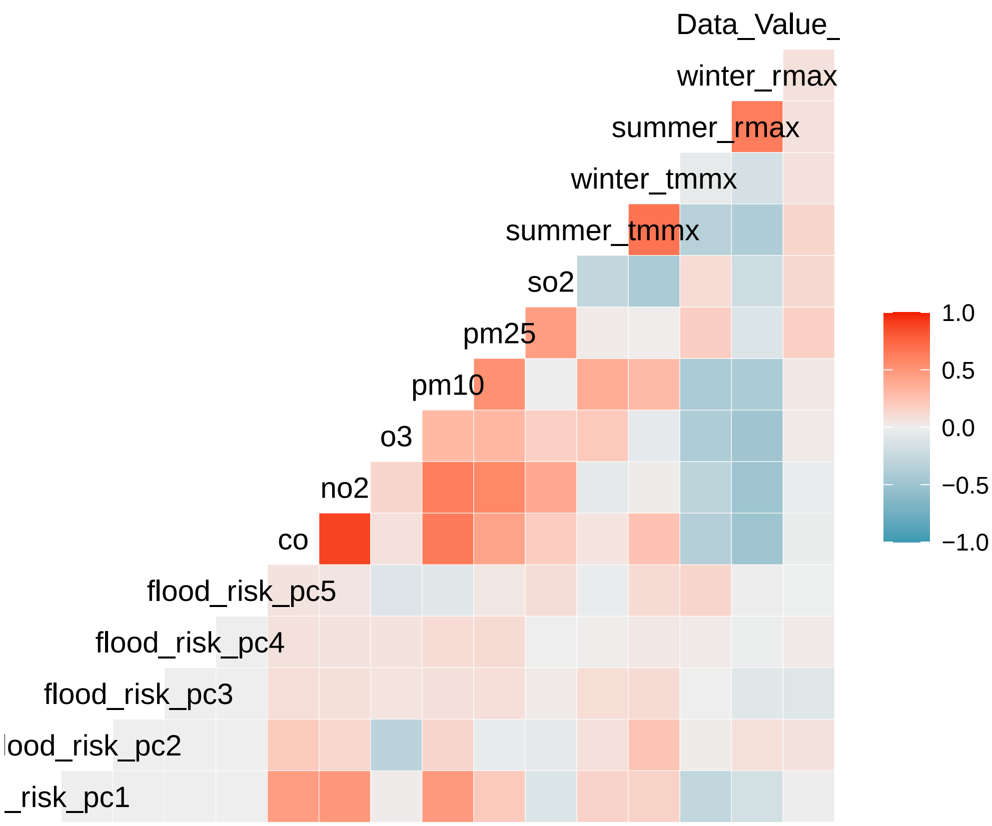
```
fhs_model_df <- readRDS(here("intermediary_data/fhs_model_df_all_census_tract_pc.rds"))
```

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

Excluding variables in collin_var_names

```
fr_index <- 19:23

climate_var_idx <- c(fr_index, 40:49)
```

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




```r
climate_cor <- cor(fhs_model_df[complete.cases(fhs_model_df[, c(climate_var_idx, ncol(fhs_model_df) - 3

climate_cor[nrow(climate_cor), ] # correlation with CHD
```

```
## flood_risk_pc1 flood_risk_pc2 flood_risk_pc3 flood_risk_pc4 flood_risk_pc5
##    -0.23118564     0.09085834    -0.11083912    -0.01721543    -0.01371888
##             co            no2             o3           pm10           pm25
##    -0.26411360    -0.31124317    -0.11232849    -0.17635941    -0.07814542
##            so2    summer_tmmx    winter_tmmx    summer_rmax    winter_rmax
##     0.05324431     0.10088375     0.04387705     0.18700816     0.20222844
## Data_Value_CHD
##     1.00000000
```

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

```r
diag(climate_cor) <- NA

summary(climate_cor)
```

```
##  flood_risk_pc1     flood_risk_pc2      flood_risk_pc3
##  Min.   :-0.28267   Min.   :-0.318203   Min.   :-0.1108391
##  1st Qu.:-0.06116   1st Qu.:-0.006856   1st Qu.:-0.0002654
##  Median : 0.01360   Median : 0.015618   Median : 0.0215093
##  Mean   : 0.07910   Mean   : 0.038553   Mean   : 0.0303619
##  3rd Qu.: 0.18469   3rd Qu.: 0.112164   3rd Qu.: 0.0906440
##  Max.   : 0.49194   Max.   : 0.242600   Max.   : 0.1082687
##  NA's   :1          NA's   :1           NA's   :1
##  flood_risk_pc4       flood_risk_pc5            co                 no2
##  Min.   :-0.0172154   Min.   :-0.102194   Min.   :-0.44677   Min.   :-0.46588
```

```
##  1st Qu.: 0.0006816   1st Qu.:-0.008280   1st Qu.: 0.05030   1st Qu.:-0.03119
##  Median : 0.0056590   Median : 0.005621   Median : 0.09457   Median : 0.08919
##  Mean   : 0.0312939   Mean   : 0.019695   Mean   : 0.15276   Mean   : 0.15309
##  3rd Qu.: 0.0680582   3rd Qu.: 0.057160   3rd Qu.: 0.33575   3rd Qu.: 0.41635
##  Max.   : 0.1112367   Max.   : 0.145526   Max.   : 0.88323   Max.   : 0.88323
##  NA's   :1            NA's   :1           NA's   :1          NA's   :1
##        o3                 pm10                pm25                so2
##  Min.   :-0.509809   Min.   :-0.36175   Min.   :-0.12743   Min.   :-0.41123
##  1st Qu.:-0.107261   1st Qu.:-0.04249   1st Qu.: 0.02787   1st Qu.:-0.08752
##  Median : 0.064053   Median : 0.13787   Median : 0.11124   Median : 0.02151
##  Mean   :-0.004311   Mean   : 0.16958   Mean   : 0.18789   Mean   : 0.02854
##  3rd Qu.: 0.176250   3rd Qu.: 0.41856   3rd Qu.: 0.38000   3rd Qu.: 0.12660
##  Max.   : 0.347448   Max.   : 0.62556   Max.   : 0.57457   Max.   : 0.46857
##  NA's   :1           NA's   :1          NA's   :1          NA's   :1
##   summer_tmmx        winter_tmmx          summer_rmax         winter_rmax
##  Min.   :-0.35431   Min.   :-0.411235   Min.   :-0.398959   Min.   :-0.50981
##  1st Qu.:-0.05138   1st Qu.: 0.001599   1st Qu.:-0.285627   1st Qu.:-0.35803
##  Median : 0.04742   Median : 0.043877   Median :-0.002361   Median :-0.14623
##  Mean   : 0.05047   Mean   : 0.085709   Mean   :-0.043919   Mean   :-0.13138
##  3rd Qu.: 0.12706   3rd Qu.: 0.202201   3rd Qu.: 0.129579   3rd Qu.:-0.00342
##  Max.   : 0.69237   Max.   : 0.692370   Max.   : 0.596532   Max.   : 0.59653
##  NA's   :1          NA's   :1           NA's   :1           NA's   :1
##  Data_Value_CHD
##  Min.   :-0.31124
##  1st Qu.:-0.14434
##  Median :-0.01722
##  Mean   :-0.04247
##  3rd Qu.: 0.07205
##  Max.   : 0.20223
##  NA's   :1
```

Climate variables other than flood risk are not too correlated.

Using VIF to exclude variables

```r
X <- fhs_model_df[, 19:(ncol(fhs_model_df) - 4)]



X           <- scale(X) # Scale covariates

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

```r
vif(X)
```

```
##              Variables       VIF
## 1        flood_risk_pc1  1.992709
## 2        flood_risk_pc2  1.399009
## 3        flood_risk_pc3  1.115523
## 4        flood_risk_pc4  1.058498
## 5        flood_risk_pc5  1.105053
## 6                EP_POV  3.675117
## 7              EP_UNEMP  1.950067
## 8                EP_PCI  3.026982
```

```
## 9            EP_NOHSDP  5.542745
## 10            EP_AGE65  2.445972
## 11            EP_AGE17  2.852665
## 12           EP_DISABL  2.741188
## 13           EP_SNGPNT  2.756702
## 14           EP_MINRTY  3.944439
## 15           EP_LIMENG  3.895060
## 16            EP_MUNIT  1.946033
## 17           EP_MOBILE  1.670855
## 18            EP_CROWD  2.744679
## 19            EP_NOVEH  2.809440
## 20           EP_GROUPQ  1.450111
## 21          EP_UNINSUR  2.525256
## 22                  co  9.071307
## 23                 no2 13.521617
## 24                  o3  2.949653
## 25                pm10  3.827427
## 26                pm25  5.269781
## 27                 so2  2.614560
## 28         summer_tmmx  4.354231
## 29         winter_tmmx  5.176801
## 30         summer_rmax  3.599677
## 31         winter_rmax  3.141842
## 32 Data_Value_CSMOKING  6.234731
```

```
vifstep(X)
```

```
## 1 variables from the 32 input variables have collinearity problem:
##
## no2
##
## After excluding the collinear variables, the linear correlation coefficients ranges between:
## min correlation ( o3 ~ EP_GROUPQ ):  2.272026e-05
## max correlation ( EP_CROWD ~ EP_LIMENG ):  0.7059339
##
## ---------- VIFs of the remained variables --------
##              Variables      VIF
## 1       flood_risk_pc1 1.742477
## 2       flood_risk_pc2 1.342389
## 3       flood_risk_pc3 1.090304
## 4       flood_risk_pc4 1.042340
## 5       flood_risk_pc5 1.105922
## 6               EP_POV 3.565768
## 7             EP_UNEMP 1.934778
## 8               EP_PCI 2.907060
## 9            EP_NOHSDP 5.551178
## 10            EP_AGE65 2.312008
## 11            EP_AGE17 2.672469
## 12           EP_DISABL 2.688598
## 13           EP_SNGPNT 2.515753
## 14           EP_MINRTY 3.632616
## 15           EP_LIMENG 3.800861
## 16            EP_MUNIT 2.029194
## 17           EP_MOBILE 1.584014
## 18            EP_CROWD 2.826198
```

```
## 19           EP_NOVEH 2.858373
## 20          EP_GROUPQ 1.537550
## 21         EP_UNINSUR 2.418424
## 22                 co 4.651525
## 23                 o3 2.639130
## 24               pm10 3.718772
## 25               pm25 4.425108
## 26                so2 2.549814
## 27        summer_tmmx 4.267137
## 28        winter_tmmx 4.458669
## 29        summer_rmax 3.642509
## 30        winter_rmax 2.777570
## 31 Data_Value_CSMOKING 6.224628
```

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

```
collin_var_names <- c("no2")
```

Non-spatial modeling

```
Y <- fhs_model_df$Data_Value_CHD

X <- fhs_model_df[, 19:(ncol(fhs_model_df) - 4)]

# exclude some more variables selected by vifstep, to account for multicollinearity

collin_var_names <- c("no2")

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



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
## -9.9860 -0.4808 -0.0182  0.4573 17.7404
##
## Coefficients:
```

```
##                        Estimate Std. Error  t value Pr(>|t|)    
## (Intercept)            6.660115   0.003138 2122.297  < 2e-16 ***
## Xflood_risk_pc1       -0.027024   0.004079   -6.625 3.49e-11 ***
## Xflood_risk_pc2       -0.001688   0.003741   -0.451 0.651729    
## Xflood_risk_pc3       -0.034736   0.003400  -10.216  < 2e-16 ***
## Xflood_risk_pc4       -0.002937   0.003306   -0.888 0.374345    
## Xflood_risk_pc5       -0.020495   0.003395   -6.037 1.58e-09 ***
## XEP_POV                0.340454   0.005937   57.346  < 2e-16 ***
## XEP_UNEMP              0.015919   0.004359    3.652 0.000261 ***
## XEP_PCI               -0.025340   0.005230   -4.845 1.27e-06 ***
## XEP_NOHSDP             0.212585   0.007404   28.710  < 2e-16 ***
## XEP_AGE65              1.472273   0.004753  309.762  < 2e-16 ***
## XEP_AGE17              0.337517   0.005386   62.665  < 2e-16 ***
## XEP_DISABL             0.348234   0.005173   67.318  < 2e-16 ***
## XEP_SNGPNT            -0.101834   0.005084  -20.029  < 2e-16 ***
## XEP_MINRTY            -0.067124   0.005983  -11.219  < 2e-16 ***
## XEP_LIMENG            -0.007475   0.006225   -1.201 0.229860    
## XEP_MUNIT             -0.056694   0.004488  -12.633  < 2e-16 ***
## XEP_MOBILE             0.041904   0.003997   10.485  < 2e-16 ***
## XEP_CROWD             -0.069757   0.005343  -13.055  < 2e-16 ***
## XEP_NOVEH              0.036788   0.005579    6.594 4.31e-11 ***
## XEP_GROUPQ            -0.076350   0.003880  -19.676  < 2e-16 ***
## XEP_UNINSUR            0.151671   0.004836   31.365  < 2e-16 ***
## Xco                    0.017966   0.007075    2.539 0.011115 *  
## Xo3                   -0.060357   0.005025  -12.012  < 2e-16 ***
## Xpm10                 -0.013396   0.006204   -2.159 0.030838 *  
## Xpm25                 -0.006493   0.006439   -1.008 0.313229    
## Xso2                   0.082266   0.005180   15.882  < 2e-16 ***
## Xsummer_tmmx           0.114721   0.006607   17.363  < 2e-16 ***
## Xwinter_tmmx           0.062928   0.006596    9.541  < 2e-16 ***
## Xsummer_rmax           0.064351   0.006218   10.349  < 2e-16 ***
## Xwinter_rmax           0.078880   0.005644   13.975  < 2e-16 ***
## XData_Value_CSMOKING   0.838385   0.007549  111.059  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.8406 on 71803 degrees of freedom
##   (702 observations deleted due to missingness)
## Multiple R-squared:  0.855,  Adjusted R-squared:  0.855 
## F-statistic: 1.366e+04 on 31 and 71803 DF,  p-value: < 2.2e-16
```