Economic Output and Risk Factors

Nathan Seto

What is the relationship between health risk concentrations and economic output?

Calculate correlations between environmental exposures and farm sales (and/or other econ. variables) to understand economic linkages.

To see coefficients and significance, follow this link: [https://nseto.shinyapps.io/correlation_econoutput_environmentalrisk/].

The values used as variables of economic output are listed below. The predictor variables are all the variables in each data set that are not in the following list of variables. Variables ending in "$_operation" denote "dollars per operation".

```
## [[1]]
##  [1] "commodity_totals_sales_measured_in_$"
##  [2] "commodity_totals_sales_measured_in_$_operation"
##  [3] "crop_totals_sales_measured_in_$"
##  [4] "animal_totals_incl_products_sales_measured_in_$"
##  [5] "commodity_totals_operations_with_sales"
##  [6] "farm_sales_(less_than_2500_$)"
##  [7] "farm_sales_(2500_to_4999_$)"
##  [8] "farm_sales_(5000_to_9999_$)"
##  [9] "farm_sales_(10000_to_24999_$)"
## [10] "farm_sales_(25000_to_49999_$)"
## [11] "farm_sales_(50000_to_99999_$)"
## [12] "farm_sales_(100000_or_more_$)"
## [13] "income_farmrelated_operations_with_receipts"
## [14] "income_farmrelated_receipts_measured_in_$"
## [15] "income_net_cash_farm_of_operations_operations_with_net_income"
## [16] "income_net_cash_farm_of_operations_net_income_measured_in_$"
## [17] "income_net_cash_farm_of_operations_net_income_measured_in_$_operation"
```

Methods

I started this process by collecting separate data sets for all counties in the states covered by HICAHS from the United States Department of Agriculture agricultural census database. I then merged all these data sets into a single file. Since the variables were in rows, I pivoted the data set so the variables were made into columns. Following this, I merged the data containing natural disaster/weather variables, creating the complete data I desired to work with. See 'Processing\output_Risk.R' for specific methods. Correlation coefficients were reached by taking one predictor and calculating the Spearman correlation coefficient and p-value between the remaining variables. This was done for all 17 response variables above. These values were then entered into a table to be displayed. To create the correlation plots found below, the original data set was split into 12 data sets by the type of natural disaster/weather factor, then correlations were

calculated for each pair and graphed. The 17 response variables are identical throughout all 12 data sets. All work was completed using R and RStudio.

The Correlations
Most of the correlations found were between 0.50 and 0.76, meaning economic output has a moderate to decently strong positive correlation with environmental risk factors used in this analysis. `Commodity Total Sales Measured in Dollars` had the highest correlations, almost reaching 1. This is analogous to the assumption that agricultural production and output would be hindered by increased damaging environmental effects.

Similar results are found using Spearman's correlation coefficient.

Significance
Many correlations this analysis provided were significant ($p < 0.05$). A large number of p-values in the table are shown as zero. This means either the p-value is, in fact, zero, or very nearly zero. P-values equal to zero mean that the result is statistically impossible to occur under the null hypothesis. The number of p-values that are shown as zero is alarming, which brings to question the methods of testing I used or the structure of the data. However, I did use 'cor.test()' whose results I have no reason to question the validity of.

Correlograms
Below are a series of correlation plots between variables of economic output and different weather/natural disaster predictors. The main focus is the rectangle towards the center of the plots, more prominent in some plots than others. The disaster/weather variables are naturally correlated with each other a majority of the time. The response variables are naturally correlated with each other to some degree. This is to be expected.
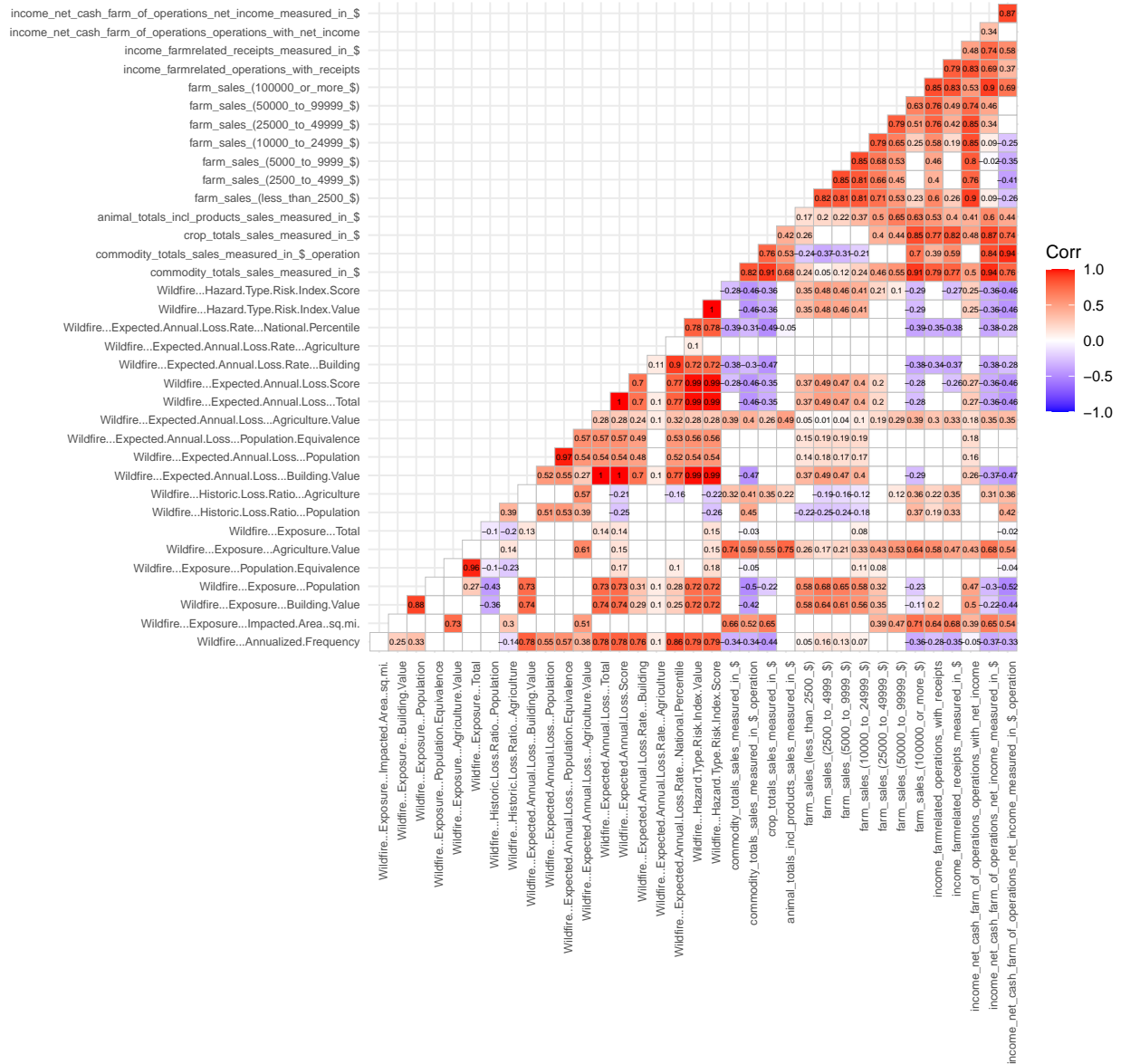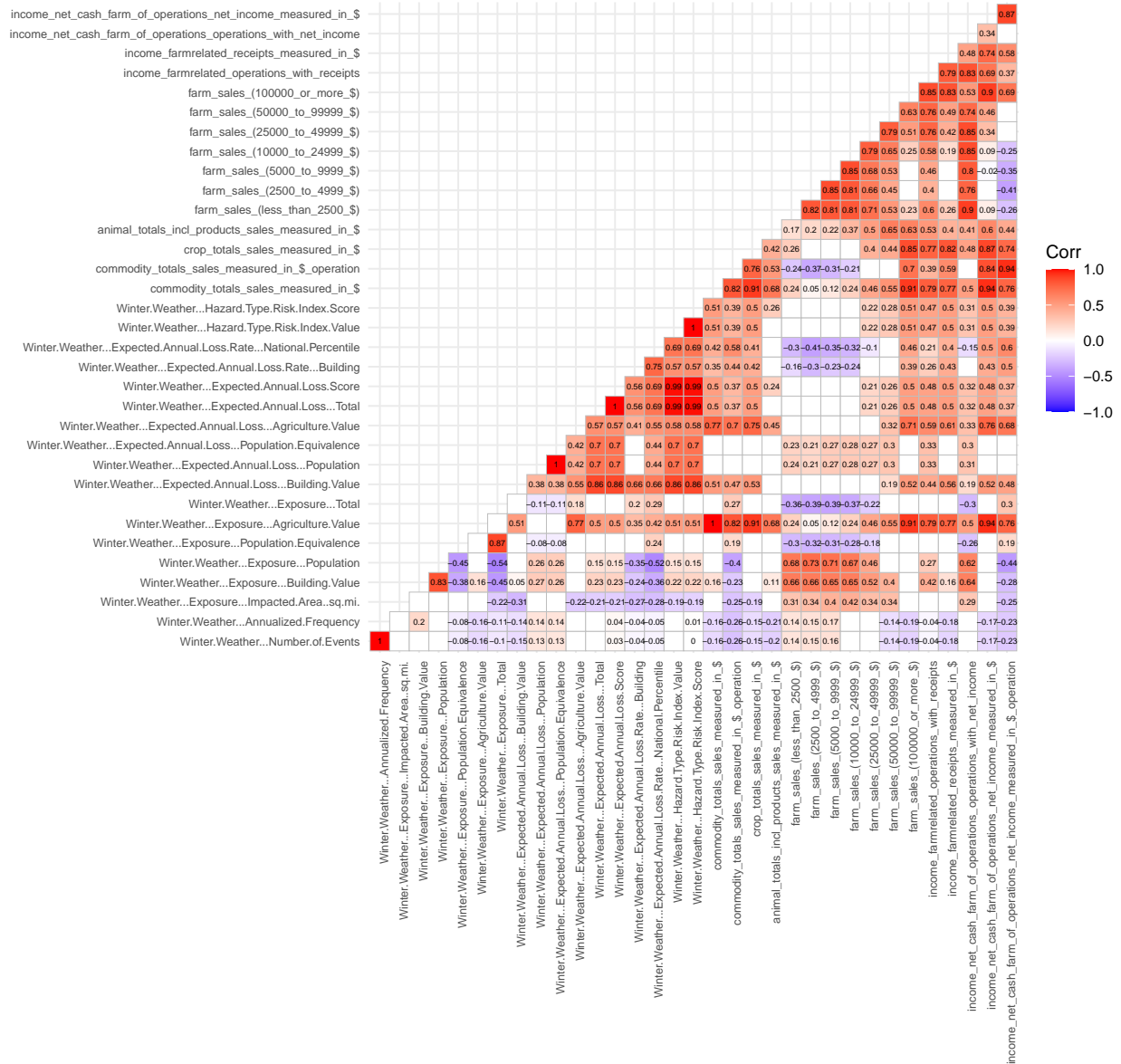
Correlation Plot: Drought
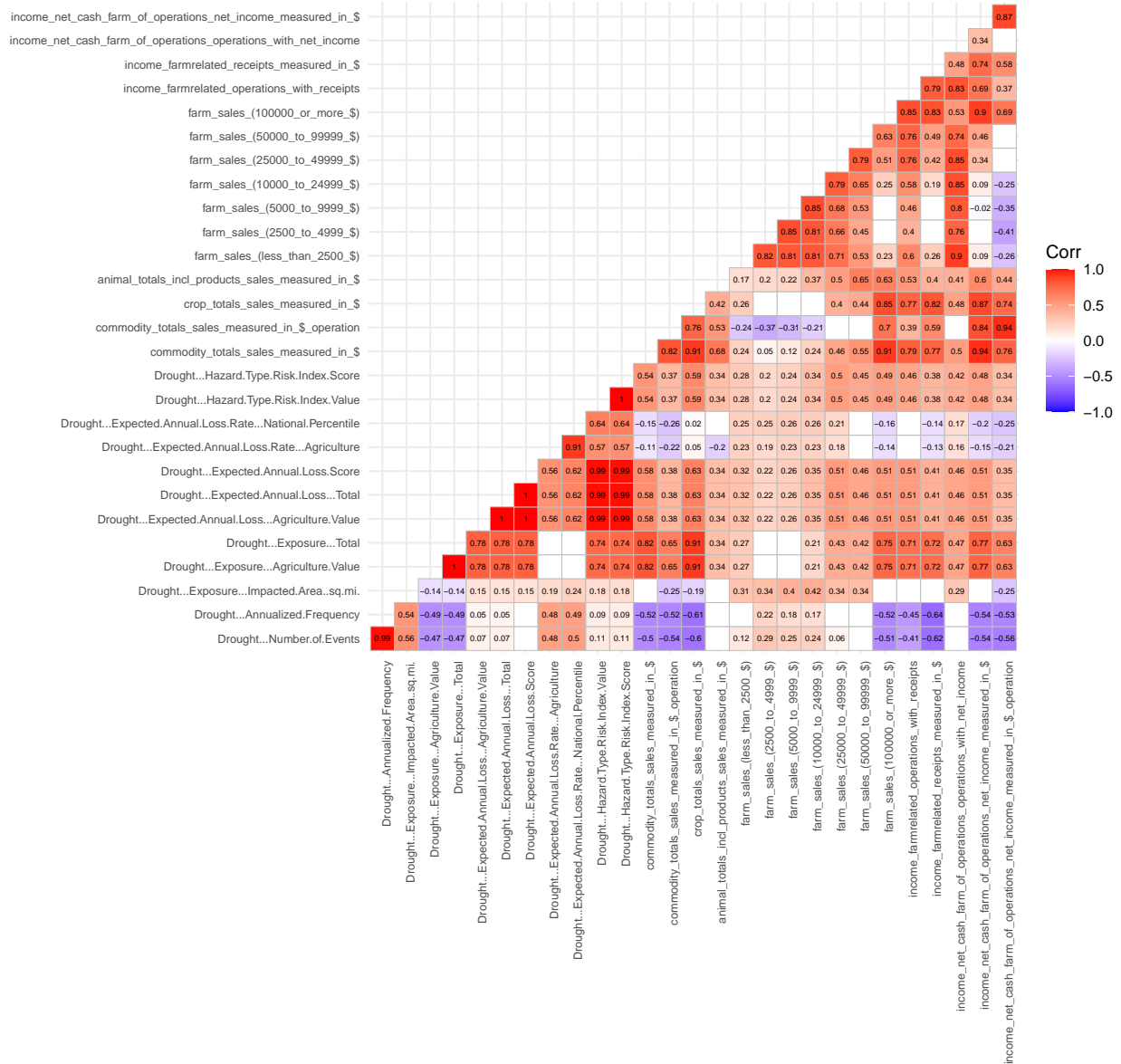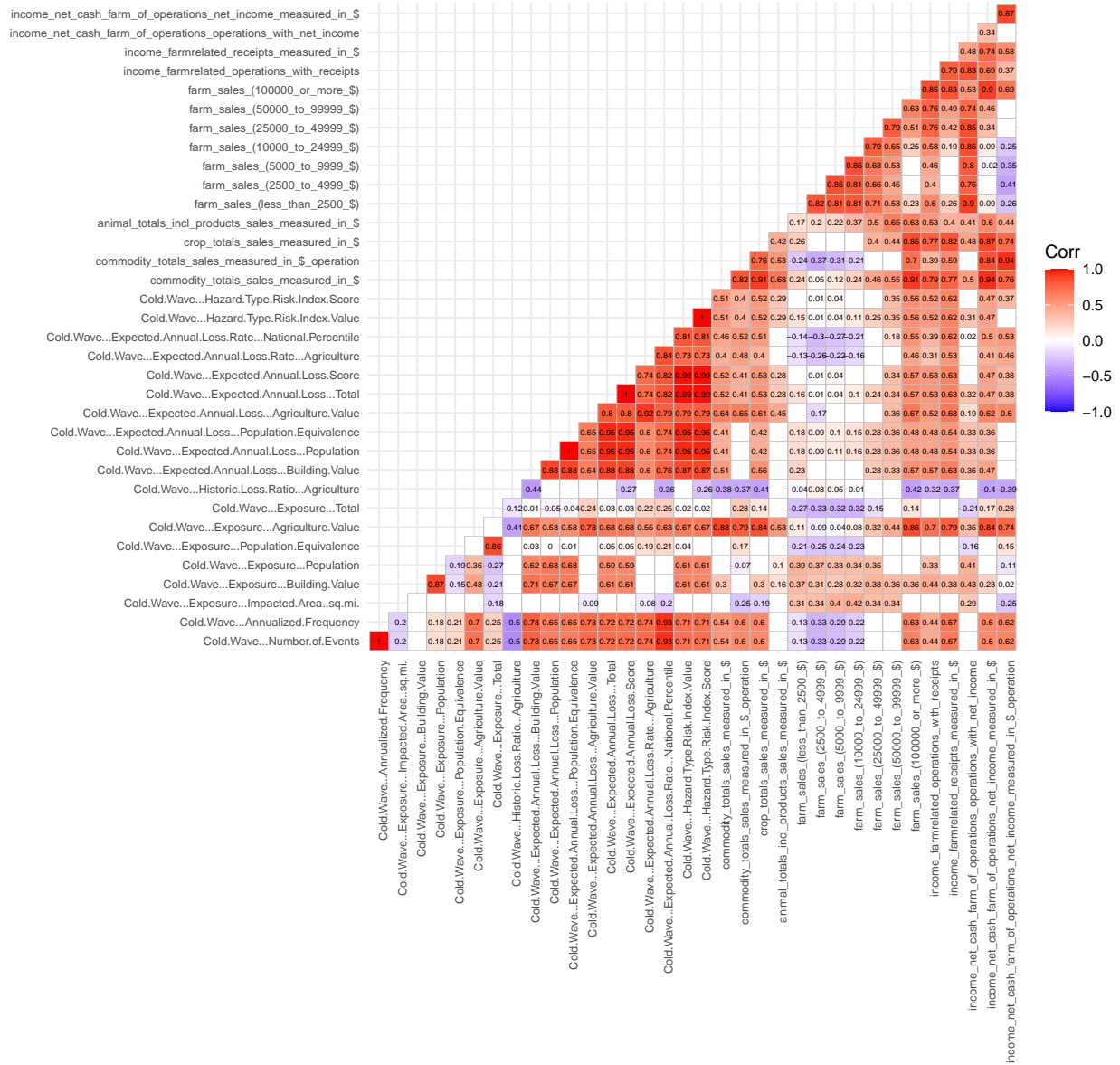
Correlation Plot: Heat Wave
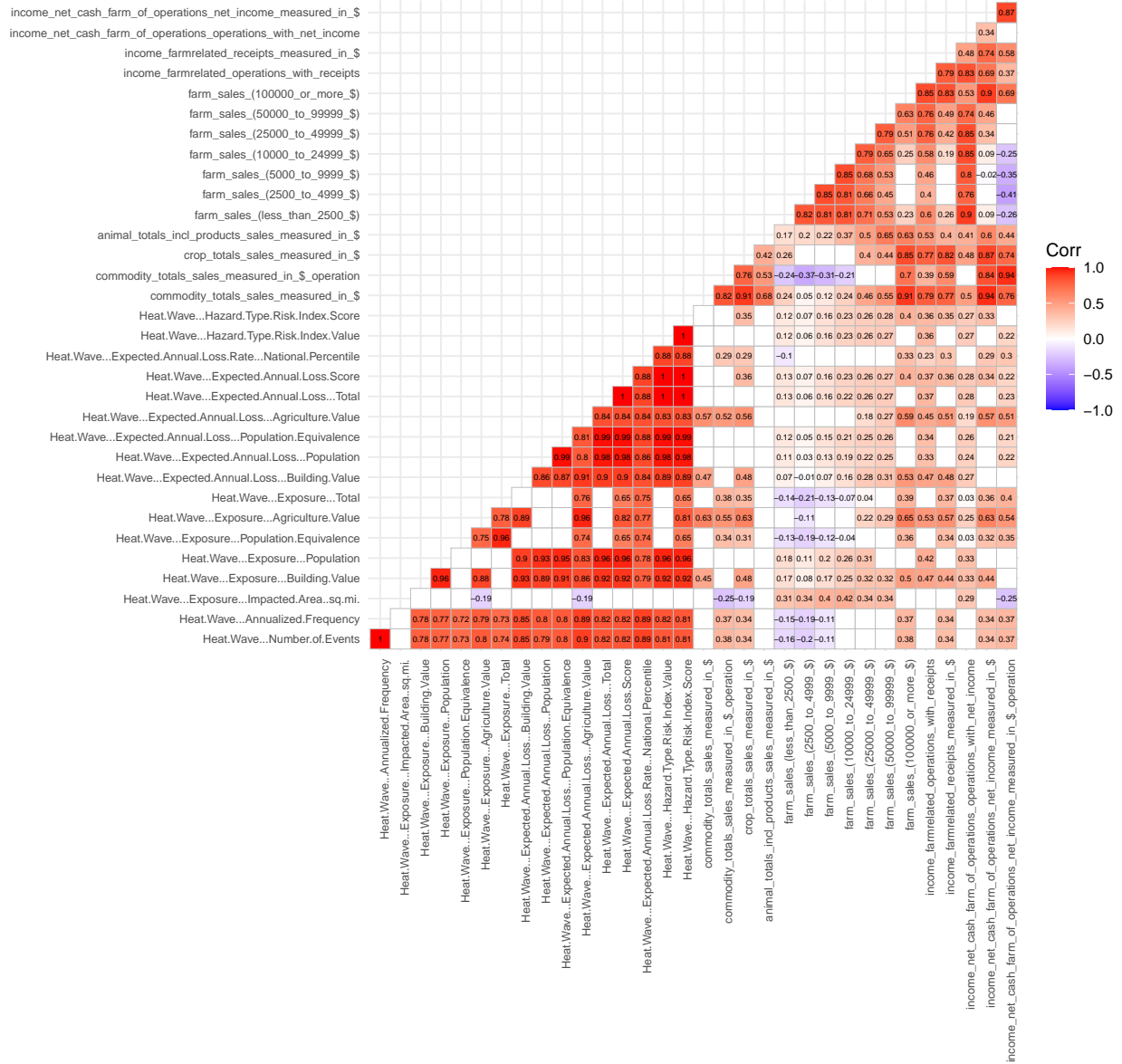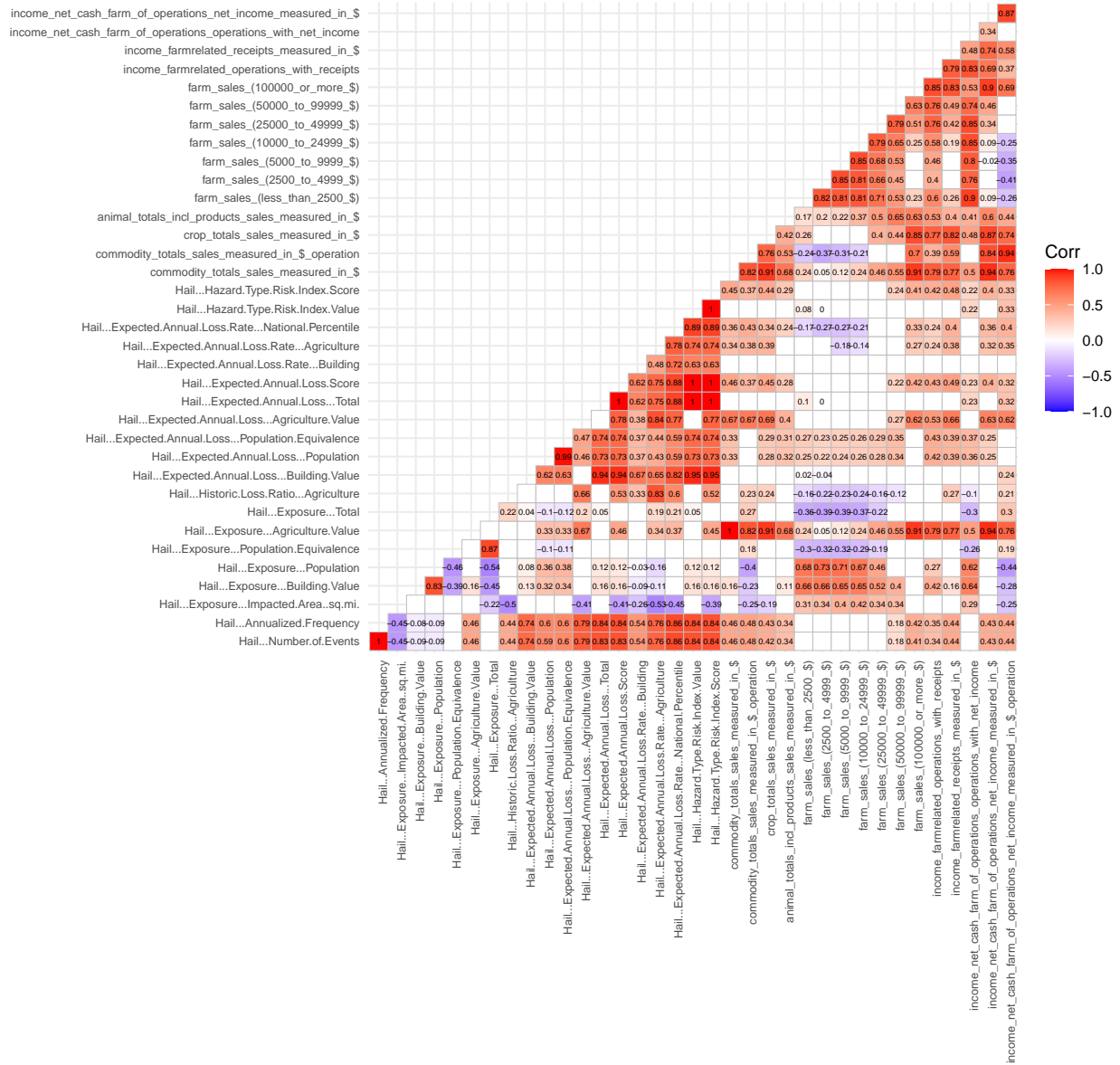
Correlation Plot: Wildfire
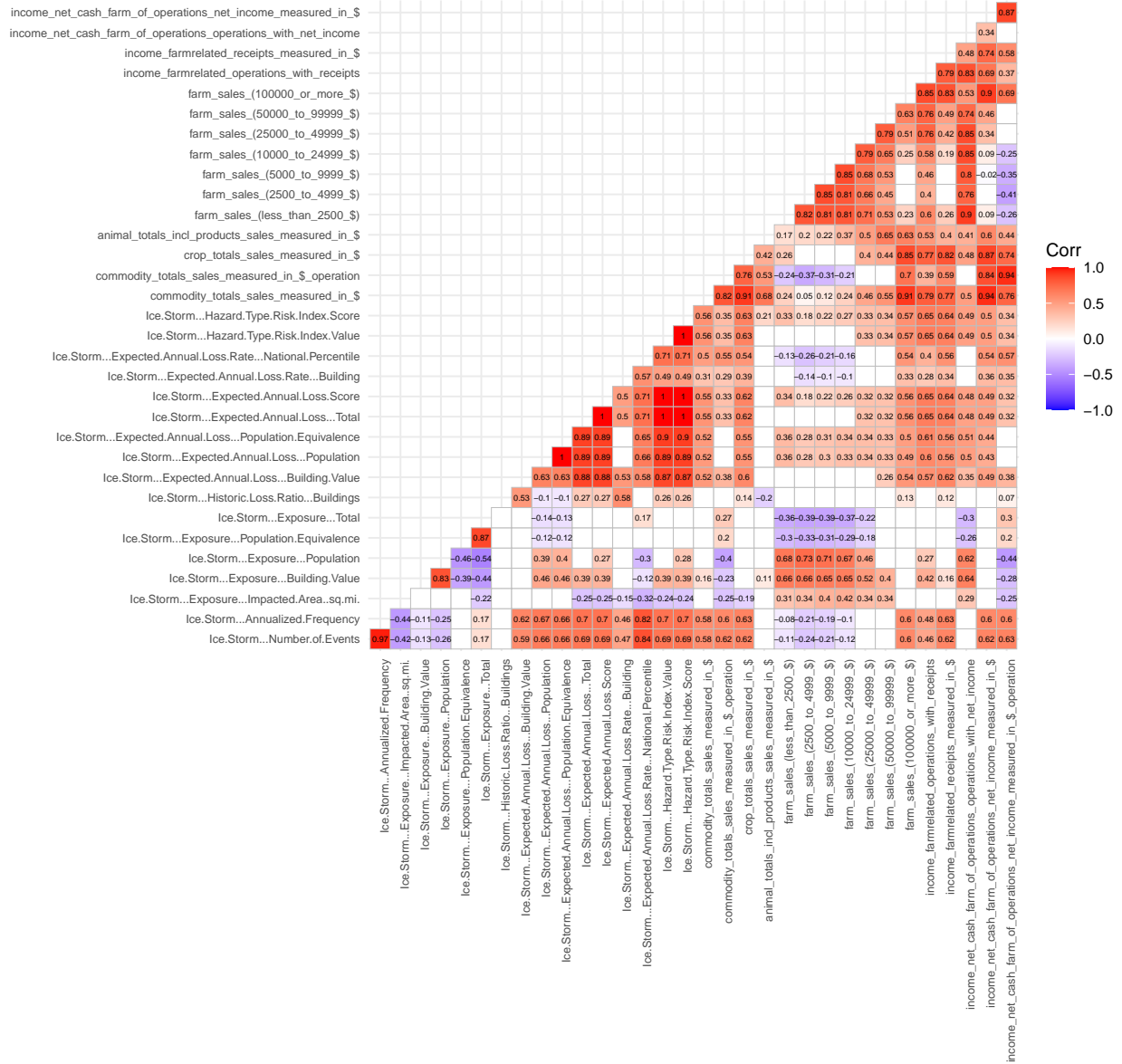
Correlation Plot: Winter Weather
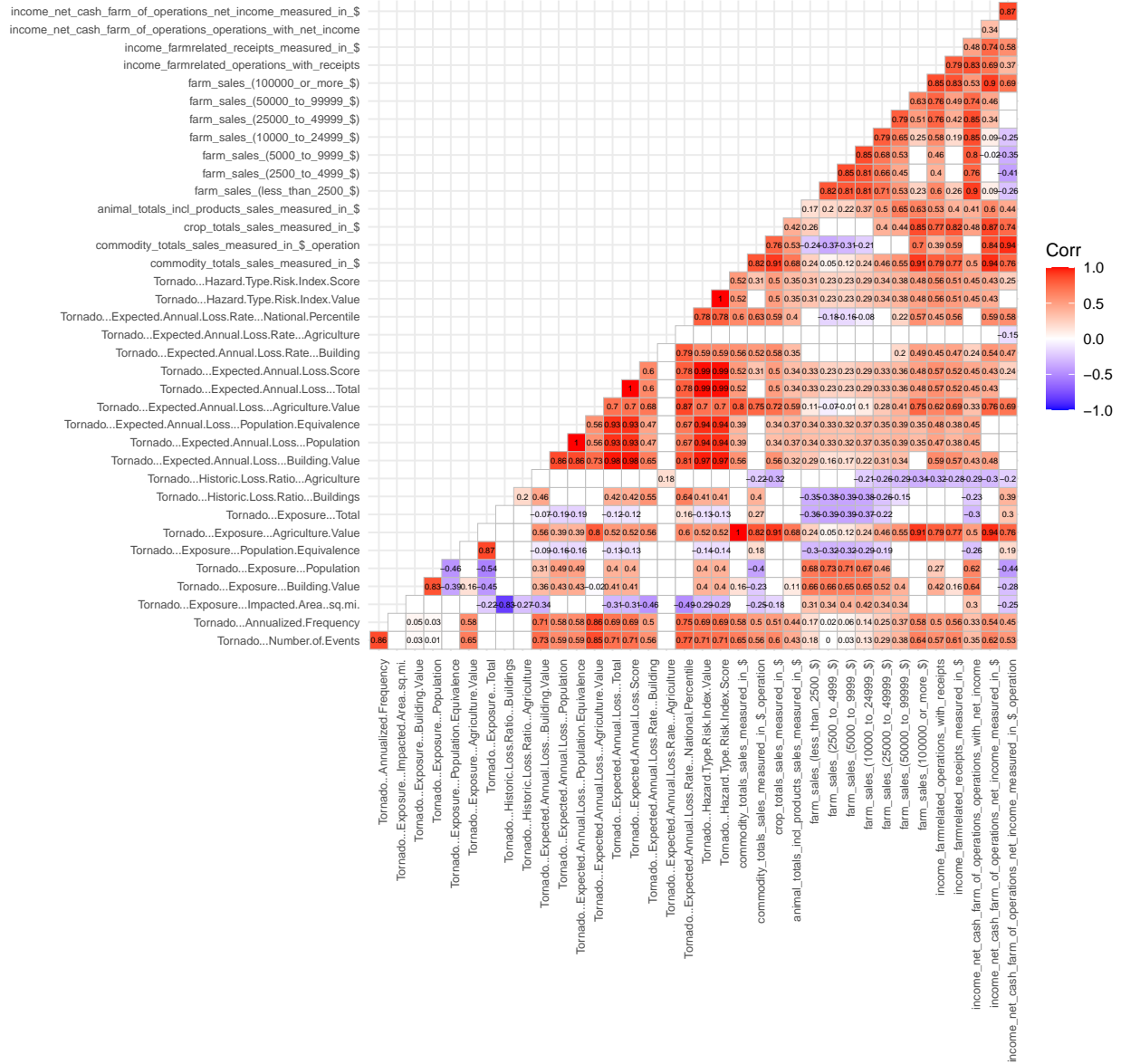
Correlation Plot: Cold Wave

Correlation Plot: Hail
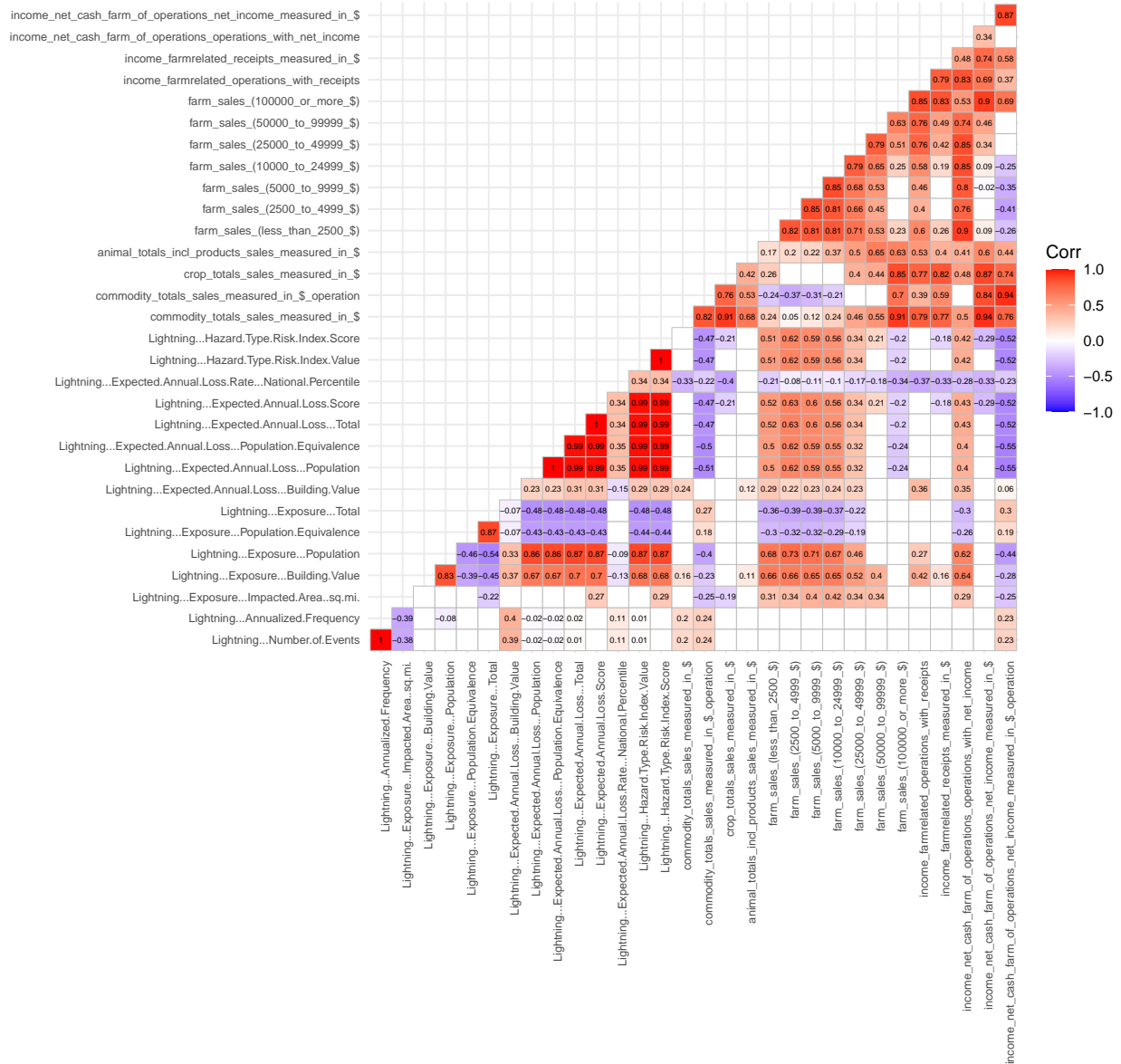
Correlation Plot: Ice Storm
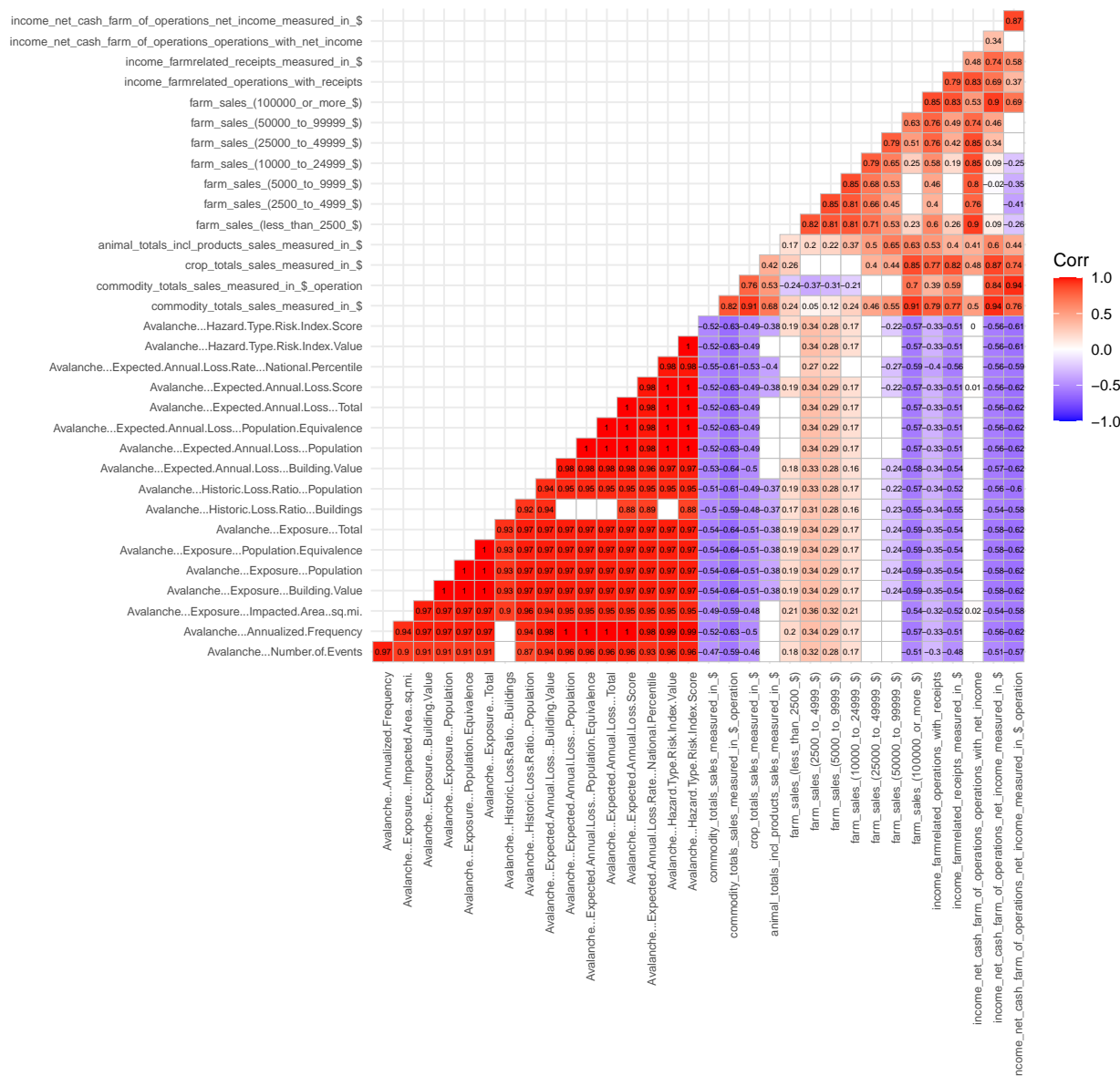
Correlation Plot: Strong Wind

Correlation Plot: Tornado

Correlation Plot: Lightning

Correlation Plot: Avalanche

Problems with Data Structure

Many of the columns are filled with NAs, but I am lead to believe this is because some factors simply do not happen in certain states/counties. This could also be due to how to data was transformed into a clean data set; further investigation is required. When processing the data prior to calculating correlations, I changed all NAs to 0, which may have caused faulty evaluation of correlations and/or significance as well. I chose to do this because many variables take the frequency or exposure from different variables, so if a county never gets any events of such then frequency, exposure, or any related factor is virtually nonexistent, and the data entry is entered as NA, and is therefore changed to zero by my data cleaning, which is why converting those value is a valid procedure. Changing NAs to zero also make it easier for R to calculate correlation coefficients. Keeping NA values can lead to the deletion of rows depending on the requirements of the algorithms used. Since each row is a different county, I want to avoid this as much as I can in order to have

the most complete data set possible (I have not removed any rows at this point). Mean/Median or K-Nearest Neighbor imputation was out of the question for this data. Doing so would have altered recorded fact and created falsehoods, adding bias that would otherwise not be present and negating the overall plausibility of the analysis. Removing rows without complete observations was not favorable for similar reasons. Since the purpose is to find correlations between variables and not related to machine learning at this time, the optimal method was converting NAs to 0 values to create complete observations.