| **Script Name** | **What Script Does** | **Inputs** | **Outputs** |
| --- | --- | --- | --- |
| SuperLearner1.R | * Uses SuperLearner function to apply a battery of machine learning classification methods to construct an optimal new predictor | * training.rda * testing.rda * listsplits.rda | * SLfit |
| create\_samplesplits\_of\_training\_data.R | * Creates a list of ten vectors with all records in SE plus random draws without replacement of all other records |  |  |
| merge\_ACS\_SWDI\_SE.R | * Merges ACS data to the merged SWDI/SE data (after getting rid of SWDI/SE records that could not be geocoded because they were in the ocean or Canada or Mexico). * Constructs explanatory variables. * Creates random uniform vector for use later in constructing CV folds. * Splits data frame into training and test sets. | * ACS\_2006\_2010\_counties\_Tracts.rda * SWDI\_SE\_geo.rda | * SWDI\_SE\_ACS.rda * training.rda * testing.rda |
| prepare\_SWDI\_SE\_for\_ACS\_merge.R  prepare\_SWDI\_SE\_for\_ACS\_merge\_remainder.R | * Creates dataframe of unique coordinates from the SWDI/SE file * Uses FCC API to create a table of coordinates and their corresponding 2010 Census Block FIPS code * Takes a long time to run; had to fill in the remainder of table after a time out. * Note that some SWDI coordinates could not be matched because they were not located on US land territory (i.e., in an ocean or in Canada or Mexico). | * SWDI\_SE\_merge\_2011\_2015.rda | * SWDI\_SE\_geo.rda |
| merge\_ACS\_county\_tract.R | * Merges tract- and county-level ACS data | * ACS\_2006\_2010\_counties.rda * ACS\_2006\_2010.rda | * ACS\_2006\_2010\_counties\_Tracts.rda |
| ACS\_county\_level.R | * Downloads 2006-2010 ACS Summary File data by county * Merges it with Census geodata by county | * ACS Summary File data via API calls from acs package * Gaz\_counties\_national.txt | * ACS\_2006\_2010\_counties.rda |
| ACS.R | * Downloads 2006-2010 ACS Summary File data by census tract * Merges it with Census geodata by census tract | * ACS Summary File data via API calls from acs package * Gaz\_tracts\_national.txt | * ACS\_2006\_2010.rda |
| assess\_merge\_SWDI\_SE\_hail\_2011\_2015\_new.R | * Merges SWDI and SE data and assesses the quality of the merge | For years 2011-2015:   * DeDuplicateHailStormEventData\_[year].rda * DeDuplicateSWDIHailData\_[year].rda | * SWDI\_SE\_merge\_2011\_2015.rda |
| prepare\_SWDI\_for\_storm\_events\_data\_merge\_2011\_2015.R | * Reads in raw SWDI hail data * Limits data to lower 48 states * Rounds coordinates to the nearest ¼ degree * Deduplicates data by date/coordinates keeping max/min for SEVPROB and MAXSIZE | For years 2011-2015:   * Nexrad\_Hail\_[year].rda | For years 2011-2015:   * DeDuplicateSWDIHailData\_[year].rda |
| prepare\_storm\_events\_data\_for\_SWDI\_merge\_2011\_2015.R | * Extracts hail data from Storm Events data * Rounded coordinates to nearest ¼ degree * Converted times (hence dates) from local standard time to GMT * Used beginning and ending coordinates to develop records reflecting full path of hail storms * Deduplicated by date/coordinates keeping max/min values of hail size, economic and health damages | For years 2011-2015:   * merged\_SE\_[year].rda | For years 2011-2015:   * DeDuplicateHailStormEventData\_[year].rda |
| StormEvents\_2008\_2015\_merge\_location\_details.R | * Downloads and unzips Storm Event details and location files and merges them by Episode and Event ID | <http://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/> | For years 2008-2015:   * merged\_SE\_[year].rda |
| NexradHail\_2007\_2015.R | * Downloads SWDI data on hail events | See CDS tutorial on hail data. | For years 2007-2015:   * Nexrad\_Hail\_[year].rda |