Strorm Data Affect Analysis: What are the Public Economics and Health Effects?

grldsndrs

```
## [1] "Fri Jan 29 16:36:21 2016"
```

Data Processing: Load Data via *"ProjectTemplate"*

```
> ldData <- process.loadData()
```

Loading project configuration

Autoloading helper functions

 Running helper script: controllerGeneratorHelp.R

 Running helper script: helpers.R

 Running helper script: modelGeneratorHelp.R

 Running helper script: viewGeneratorHelp.R

Autoloading packages

 Loading package: reshape

 Loading package: plyr

 Loading package: ggplot2

 Loading package: stringr

 Loading package: lubridate

 Loading package: dplyr

 Loading package: graphics

 Loading package: rmarkdown

 Loading package: knitr

Autoloading cache

```
Autoloading data

Converting data.frames to data.tables

Munging data

 Running preprocessing script: 00-modelQueryHelp.R

 Running preprocessing script: 00-transformationsHelp.R

 Running preprocessing script: 01-A.R


$repdata_data_StormData.csv.bz2
$repdata_data_StormData.csv.bz2$tag
[1] "Summary"


$repdata_data_StormData.csv.bz2$out
    STATE__                 BGN_DATE                BGN_TIME
 Min.   : 1.0   5/25/2011 0:00:00:  1202   12:00:00 AM: 10163
 1st Qu.:19.0   4/27/2011 0:00:00:  1193   06:00:00 PM:  7350
 Median :30.0   6/9/2011 0:00:00 :  1030   04:00:00 PM:  7261
 Mean   :31.2   5/30/2004 0:00:00:  1016   05:00:00 PM:  6891
 3rd Qu.:45.0   4/4/2011 0:00:00 :  1009   12:00:00 PM:  6703
 Max.   :95.0   4/2/2006 0:00:00 :   981   03:00:00 PM:  6700
                (Other)          :895866   (Other)    :857229
   TIME_ZONE          COUNTY          COUNTYNAME         STATE
 CST    :547493   Min.   :  0.0   JEFFERSON :  7840   TX     : 83728
 EST    :245558   1st Qu.: 31.0   WASHINGTON:  7603   KS     : 53440
 MST    : 68390   Median : 75.0   JACKSON   :  6660   OK     : 46802
 PST    : 28302   Mean   :100.6   FRANKLIN  :  6256   MO     : 35648
 AST    :  6360   3rd Qu.:131.0   LINCOLN   :  5937   IA     : 31069
 HST    :  2563   Max.   :873.0   MADISON   :  5632   NE     : 30271
 (Other):  3631                   (Other)   :862369   (Other):621339
            EVTYPE          BGN_RANGE            BGN_AZI
 HAIL             :288661   Min.   :   0.000            :547332
 TSTM WIND        :219940   1st Qu.:   0.000   N        : 86752
 THUNDERSTORM WIND: 82563   Median :   0.000   W        : 38446
 TORNADO          : 60652   Mean   :   1.484   S        : 37558
 FLASH FLOOD      : 54277   3rd Qu.:   1.000   E        : 33178
 FLOOD            : 25326   Max.   :3749.000   NW       : 24041
 (Other)          :170878                      (Other):134990
        BGN_LOCATI                END_DATE               END_TIME
            :287743                     :243411                :238978
 COUNTYWIDE   : 19680   4/27/2011 0:00:00:  1214   06:00:00 PM:  9802
 Countywide   :   993   5/25/2011 0:00:00:  1196   05:00:00 PM:  8314
 SPRINGFIELD  :   843   6/9/2011 0:00:00 :  1021   04:00:00 PM:  8104
 SOUTH PORTION:   810   4/4/2011 0:00:00 :  1007   12:00:00 PM:  7483
 NORTH PORTION:   784   5/30/2004 0:00:00:   998   11:59:00 PM:  7184
 (Other)      :591444   (Other)          :653450   (Other)    :622432
   COUNTY_END COUNTYENDN      END_RANGE            END_AZI
 Min.   :0    Mode:logical   Min.   :  0.0000            :724837
 1st Qu.:0    NA's:902297    1st Qu.:  0.0000   N        : 28082
```

```
Median :0              Median :  0.0000   S      : 22510
Mean   :0              Mean   :  0.9862   W      : 20119
3rd Qu.:0              3rd Qu.:  0.0000   E      : 20047
Max.   :0              Max.   :925.0000   NE     : 14606
                                          (Other): 72096
         END_LOCATI         LENGTH              WIDTH
                :499225  Min.   :  0.0000  Min.   :   0.000
COUNTYWIDE      : 19731  1st Qu.:  0.0000  1st Qu.:   0.000
SOUTH PORTION   :   833  Median :  0.0000  Median :   0.000
NORTH PORTION   :   780  Mean   :  0.2301  Mean   :   7.503
CENTRAL PORTION :   617  3rd Qu.:  0.0000  3rd Qu.:   0.000
SPRINGFIELD     :   575  Max.   :2315.0000 Max.   :4400.000
(Other)         :380536
       F               MAG            FATALITIES           INJURIES
Min.   :0.0      Min.   :    0.0  Min.   :  0.0000  Min.   :   0.0000
1st Qu.:0.0      1st Qu.:    0.0  1st Qu.:  0.0000  1st Qu.:   0.0000
Median :1.0      Median :   50.0  Median :  0.0000  Median :   0.0000
Mean   :0.9      Mean   :   46.9  Mean   :  0.0168  Mean   :   0.1557
3rd Qu.:1.0      3rd Qu.:   75.0  3rd Qu.:  0.0000  3rd Qu.:   0.0000
Max.   :5.0      Max.   :22000.0  Max.   :583.0000  Max.   :1700.0000
NA's   :843563
    PROPDMG           PROPDMGEXP         CROPDMG           CROPDMGEXP
Min.   :   0.00             :465934  Min.   :  0.000           :618413
1st Qu.:   0.00   K    :424665  1st Qu.:  0.000   K     :281832
Median :   0.00   M    : 11330  Median :  0.000   M     :  1994
Mean   :  12.06   0    :   216  Mean   :  1.527   k     :    21
3rd Qu.:   0.50   B    :    40  3rd Qu.:  0.000   0     :    19
Max.   :5000.00   5    :    28  Max.   :990.000   B     :     9
                  (Other):    84               (Other):     9
      WFO                               STATEOFFIC
         :142069                            :248769
OUN     : 17393   TEXAS, North                      : 12193
JAN     : 13889   ARKANSAS, Central and North Central: 11738
LWX     : 13174   IOWA, Central                     : 11345
PHI     : 12551   KANSAS, Southwest                 : 11212
TSA     : 12483   GEORGIA, North and Central        : 11120
(Other):690738   (Other)                            :595920




GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA
JEFFERSON - JEFFERSON
MADISON - MADISON
(Other)
    LATITUDE       LONGITUDE       LATITUDE_E       LONGITUDE_
Min.   :   0   Min.   :-14451  Min.   :   0   Min.   :-14455
1st Qu.:2802   1st Qu.: 7247  1st Qu.:   0   1st Qu.:     0
Median :3540   Median : 8707  Median :   0   Median :     0
Mean   :2875   Mean   : 6940  Mean   :1452   Mean   :  3509
3rd Qu.:4019   3rd Qu.: 9605  3rd Qu.:3549   3rd Qu.:  8735
Max.   :9706   Max.   : 17124 Max.   :9706   Max.   :106220
NA's   :47                    NA's   :40
                                  REMARKS          REFNUM
```

```
                                                   :287433    Min.   :     1
                                                   : 24013    1st Qu.:225575
 Trees down.\n                                      :  1110    Median :451149
 Several trees were blown down.\n                   :   568    Mean   :451149
 Trees were downed.\n                               :   446    3rd Qu.:676723
 Large trees and power lines were blown down.\n:    432    Max.   :902297
 (Other)                                            :588295
```

`$repdata_data_StormData.csv.bz2`
`$repdata_data_StormData.csv.bz2$tag`
`[1] "Structure"`

`$repdata_data_StormData.csv.bz2$out`
```
 [1] "Classes 'data.table' and 'data.frame':\t902297 obs. of  37 variables:"
 [2] " $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ..."
 [3] " $ BGN_DATE  : Factor w/ 16335 levels \"1/1/1966 0:00:00\",..: 6523 6523 4242 11116 2224 2224 226
 [4] " $ BGN_TIME  : Factor w/ 3608 levels \"00:00:00 AM\",..: 272 287 2705 1683 2584 3186 242 1683 318
 [5] " $ TIME_ZONE : Factor w/ 22 levels \"ADT\",\"AKS\",\"AST\",..: 7 7 7 7 7 7 7 7 7 7 ..."
 [6] " $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ..."
 [7] " $ COUNTYNAME: Factor w/ 29601 levels \"\",\"5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI\",..:
 [8] " $ STATE     : Factor w/ 72 levels \"AK\",\"AL\",\"AM\",..: 2 2 2 2 2 2 2 2 2 2 ..."
 [9] " $ EVTYPE    : Factor w/ 985 levels \"   HIGH SURF ADVISORY\",..: 834 834 834 834 834 834 834 834
[10] " $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ..."
[11] " $ BGN_AZI   : Factor w/ 35 levels \"\",\"  N\",\" NW\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[12] " $ BGN_LOCATI: Factor w/ 54429 levels \"\",\"- 1 N Albion\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[13] " $ END_DATE  : Factor w/ 6663 levels \"\",\"1/1/1993 0:00:00\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[14] " $ END_TIME  : Factor w/ 3647 levels \"\",\" 0900CST\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[15] " $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ..."
[16] " $ COUNTYENDN: logi  NA NA NA NA NA NA ..."
[17] " $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ..."
[18] " $ END_AZI   : Factor w/ 24 levels \"\",\"E\",\"ENE\",\"ESE\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[19] " $ END_LOCATI: Factor w/ 34506 levels \"\",\"- .5 NNW\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[20] " $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ..."
[21] " $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ..."
[22] " $ F         : int  3 2 2 2 2 2 2 1 3 3 ..."
[23] " $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ..."
[24] " $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ..."
[25] " $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ..."
[26] " $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ..."
[27] " $ PROPDMGEXP: Factor w/ 19 levels \"\",\"-\",\"?\",\"+\",..: 17 17 17 17 17 17 17 17 17 17 ..."
[28] " $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ..."
[29] " $ CROPDMGEXP: Factor w/ 9 levels \"\",\"?\",\"0\",\"2\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[30] " $ WFO       : Factor w/ 542 levels \"\",\" CI\",\"$AC\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[31] " $ STATEOFFIC: Factor w/ 250 levels \"\",\"ALABAMA, Central\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[32] " $ ZONENAMES : Factor w/ 25112 levels \"\",\"
[33] " $ LATITUDE  : num  3040 3042 3340 3458 3412 ..."
[34] " $ LONGITUDE : num  8812 8755 8742 8626 8642 ..."
[35] " $ LATITUDE_E: num  3051 0 0 0 0 ..."
[36] " $ LONGITUDE_: num  8806 0 0 0 0 ..."
[37] " $ REMARKS   : Factor w/ 436781 levels \"\",\"-2 at Deer Park\\n\",..: 1 1 1 1 1 1 1 1 1 1 ..."
[38] " $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ..."
[39] " - attr(*, \".internal.selfref\")=<externalptr> "
```

Based on the summary above I will use the *make.names* funcion to tidy up the

4

names

```
$repdata_data_StormData.csv.bz2
$repdata_data_StormData.csv.bz2$tag
[1] "Clean Names"

$repdata_data_StormData.csv.bz2$out
 [1] "STATE__"     "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
 [6] "COUNTYNAME"  "STATE"      "EVTYPE"     "BGN_RANGE"   "BGN_AZI"
[11] "BGN_LOCATI"  "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
[16] "END_RANGE"   "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
[21] "F"           "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
[26] "PROPDMGEXP"  "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
[31] "ZONENAMES"   "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
[36] "REMARKS"     "REFNUM"
```

There are very few missing values in the data set. They are in the **LATTITUDE** fields. I will leave these values missing, since latitude can be obtained from the state and county fields if neccessary.

```
> library(ggplot2)
> qplot(Sepal.Length, Petal.Length, data = iris, color = Species)
```