Module 3 Assignment

Manisha Meka

4/25/2021

#Move this into a good local directory for your current working directory and read it in to R using read_csv from the readr package.

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
Stormdetails <- read_csv ("StormEvents_details-ftp_v1.0_d1992_c20170717.csv")
```

```
##
## -- Column specification ----------------------------------------------------
## cols(
##    .default = col_double(),
##    EPISODE_ID = col_logical(),
##    STATE = col_character(),
##    MONTH_NAME = col_character(),
##    EVENT_TYPE = col_character(),
##    CZ_TYPE = col_character(),
##    CZ_NAME = col_character(),
##    WFO = col_character(),
##    BEGIN_DATE_TIME = col_character(),
##    CZ_TIMEZONE = col_character(),
##    END_DATE_TIME = col_character(),
##    DAMAGE_PROPERTY = col_character(),
##    SOURCE = col_logical(),
##    MAGNITUDE_TYPE = col_logical(),
##    FLOOD_CAUSE = col_logical(),
##    CATEGORY = col_logical(),
##    TOR_F_SCALE = col_character(),
##    TOR_OTHER_WFO = col_logical(),
```

```
##    TOR_OTHER_CZ_STATE = col_logical(),
##    TOR_OTHER_CZ_FIPS = col_logical(),
##    TOR_OTHER_CZ_NAME = col_logical()
##    # ... with 7 more columns
## )
## i Use `spec()` for the full column specifications.
```

Stormdetails

```
## # A tibble: 13,534 x 51
##     BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME END_YEARMONTH END_DAY END_TIME
##               <dbl>     <dbl>      <dbl>         <dbl>   <dbl>    <dbl>
## 1            199206        24       1511        199206      24     1511
## 2            199206        24       1827        199206      24     1827
## 3            199206        24       1943        199206      24     1943
## 4            199206        25       1950        199206      25     1950
## 5            199206        26       1251        199206      26     1251
## 6            199206        26       1840        199206      26     1840
## 7            199206        26       1840        199206      26     1840
## 8            199206        26       1843        199206      26     1843
## 9            199206        26       2010        199206      26     2010
## 10           199206        27       1428        199206      27     1428
## # ... with 13,524 more rows, and 45 more variables: EPISODE_ID <lgl>,
## #   EVENT_ID <dbl>, STATE <chr>, STATE_FIPS <dbl>, YEAR <dbl>,
## #   MONTH_NAME <chr>, EVENT_TYPE <chr>, CZ_TYPE <chr>, CZ_FIPS <dbl>,
## #   CZ_NAME <chr>, WFO <chr>, BEGIN_DATE_TIME <chr>, CZ_TIMEZONE <chr>,
## #   END_DATE_TIME <chr>, INJURIES_DIRECT <dbl>, INJURIES_INDIRECT <dbl>,
## #   DEATHS_DIRECT <dbl>, DEATHS_INDIRECT <dbl>, DAMAGE_PROPERTY <chr>,
## #   DAMAGE_CROPS <dbl>, SOURCE <lgl>, MAGNITUDE <dbl>, MAGNITUDE_TYPE <lgl>,
## #   FLOOD_CAUSE <lgl>, CATEGORY <lgl>, TOR_F_SCALE <chr>, TOR_LENGTH <dbl>,
## #   TOR_WIDTH <dbl>, TOR_OTHER_WFO <lgl>, TOR_OTHER_CZ_STATE <lgl>,
## #   TOR_OTHER_CZ_FIPS <lgl>, TOR_OTHER_CZ_NAME <lgl>, BEGIN_RANGE <dbl>,
## #   BEGIN_AZIMUTH <lgl>, BEGIN_LOCATION <lgl>, END_RANGE <dbl>,
## #   END_AZIMUTH <lgl>, END_LOCATION <lgl>, BEGIN_LAT <dbl>, BEGIN_LON <dbl>,
## #   END_LAT <dbl>, END_LON <dbl>, EPISODE_NARRATIVE <lgl>,
## #   EVENT_NARRATIVE <lgl>, DATA_SOURCE <chr>
```

#Limit the dataframe to: the beginning and ending dates and times, the episode ID, the event ID, the state name and FIPS, the "CZ" name, type, and FIPS, the event type, the source, and the beginning latitude and longitude and ending latitude and longitude (10points)

```
myvars <- c("BEGIN_YEARMONTH","BEGIN_DAY", "BEGIN_TIME", "END_YEARMONTH","END_DAY","END_TIME","EPISODE_
StormDetails_Limit<- Stormdetails[myvars]
StormDetails_Limit
```

```
## # A tibble: 13,534 x 20
##     BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME END_YEARMONTH END_DAY END_TIME
##               <dbl>     <dbl>      <dbl>         <dbl>   <dbl>    <dbl>
## 1            199206        24       1511        199206      24     1511
## 2            199206        24       1827        199206      24     1827
## 3            199206        24       1943        199206      24     1943
## 4            199206        25       1950        199206      25     1950
## 5            199206        26       1251        199206      26     1251
```

```
## 6              199206        26        1840        199206        26        1840
## 7              199206        26        1840        199206        26        1840
## 8              199206        26        1843        199206        26        1843
## 9              199206        26        2010        199206        26        2010
## 10             199206        27        1428        199206        27        1428
## # ... with 13,524 more rows, and 14 more variables: EPISODE_ID <lgl>,
## #   EVENT_ID <dbl>, STATE <chr>, STATE_FIPS <dbl>, EVENT_TYPE <chr>,
## #   CZ_TYPE <chr>, CZ_NAME <chr>, CZ_FIPS <dbl>, CZ_NAME.1 <chr>, SOURCE <lgl>,
## #   BEGIN_LAT <dbl>, BEGIN_LON <dbl>, END_LAT <dbl>, END_LON <dbl>
```

#Convert the beginning and ending dates to a "date-time" class (there should be one column for the beginning date-time and one for the ending date-time) (5 points) #Unite Begining Date and Time

```
library(tidyr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(stringi)
newdata1 <- unite(StormDetails_Limit, BEGIN_DATE, BEGIN_YEARMONTH,BEGIN_DAY, sep="")
newdata2 <- unite(newdata1, BEGIN_DATE_TIME, BEGIN_DATE, BEGIN_TIME, sep=" ")
head(newdata2)
```

```
## # A tibble: 6 x 17
##    BEGIN_DATE_TIME END_YEARMONTH END_DAY END_TIME EPISODE_ID EVENT_ID STATE
##    <chr>                   <dbl>   <dbl>    <dbl> <lgl>         <dbl> <chr>
## 1 19920624 1511          199206      24     1511 NA          9985034 COLORADO
## 2 19920624 1827          199206      24     1827 NA          9985035 COLORADO
## 3 19920624 1943          199206      24     1943 NA          9985036 COLORADO
## 4 19920625 1950          199206      25     1950 NA          9985037 COLORADO
## 5 19920626 1251          199206      26     1251 NA          9985038 COLORADO
## 6 19920626 1840          199206      26     1840 NA          9985040 COLORADO
## # ... with 10 more variables: STATE_FIPS <dbl>, EVENT_TYPE <chr>,
## #   CZ_TYPE <chr>, CZ_NAME <chr>, CZ_FIPS <dbl>, SOURCE <lgl>, BEGIN_LAT <dbl>,
## #   BEGIN_LON <dbl>, END_LAT <dbl>, END_LON <dbl>
```

#Unite End Date and Time

```
library(tidyr)
library(lubridate)
newdata3 <- unite(newdata2, END_DATE, END_YEARMONTH,END_DAY, sep="")
newdata4 <- unite(newdata3, END_DATE_TIME, END_DATE, END_TIME, sep=" ")
head(newdata4)
```

```
## # A tibble: 6 x 15
##    BEGIN_DATE_TIME END_DATE_TIME EPISODE_ID EVENT_ID STATE  STATE_FIPS EVENT_TYPE
```

```
##    <chr>          <chr>          <lgl>          <dbl> <chr>          <dbl> <chr>
## 1 19920624 1511   19920624 1511 NA           9985034 COLOR~            8 Thunderst~
## 2 19920624 1827   19920624 1827 NA           9985035 COLOR~            8 Tornado
## 3 19920624 1943   19920624 1943 NA           9985036 COLOR~            8 Tornado
## 4 19920625 1950   19920625 1950 NA           9985037 COLOR~            8 Thunderst~
## 5 19920626 1251   19920626 1251 NA           9985038 COLOR~            8 Tornado
## 6 19920626 1840   19920626 1840 NA           9985040 COLOR~            8 Tornado
## # ... with 8 more variables: CZ_TYPE <chr>, CZ_NAME <chr>, CZ_FIPS <dbl>,
## #   SOURCE <lgl>, BEGIN_LAT <dbl>, BEGIN_LON <dbl>, END_LAT <dbl>,
## #   END_LON <dbl>
```

#Change state and county names to title case (e.g., "New Jersey" instead of "NEW JERSEY") (5 points)

```
library(stringr)
state.data <- StormDetails_Limit[, c("STATE","CZ_NAME")]
state.data$state_title_case=str_to_title(state.data$STATE)
state.data$county_title_case=str_to_title(state.data$CZ_NAME)
state.data
```

```
## # A tibble: 13,534 x 4
##    STATE    CZ_NAME     state_title_case county_title_case
##    <chr>    <chr>       <chr>            <chr>
##  1 COLORADO LARIMER     Colorado         Larimer
##  2 COLORADO EL PASO     Colorado         El Paso
##  3 COLORADO KIT CARSON Colorado          Kit Carson
##  4 COLORADO MONTROSE    Colorado         Montrose
##  5 COLORADO EL PASO     Colorado         El Paso
##  6 COLORADO WELD        Colorado         Weld
##  7 COLORADO WELD        Colorado         Weld
##  8 COLORADO MORGAN      Colorado         Morgan
##  9 COLORADO LOGAN       Colorado         Logan
## 10 COLORADO WASHINGTON Colorado          Washington
## # ... with 13,524 more rows
```

#Limit to the events listed by county FIPS (CZ_TYPE of "C") and then remove the CZ_TYPE column (5 points)

```
StormDetails_CType= filter(newdata4, CZ_TYPE == "C")
StormDetails_CType$CZ_TYPE <- NULL
head(StormDetails_CType)
```

```
## # A tibble: 6 x 14
##   BEGIN_DATE_TIME END_DATE_TIME EPISODE_ID EVENT_ID STATE   STATE_FIPS EVENT_TYPE
##   <chr>           <chr>         <lgl>         <dbl> <chr>        <dbl> <chr>
## 1 19920624 1511   19920624 1511 NA           9985034 COLOR~            8 Thunderst~
## 2 19920624 1827   19920624 1827 NA           9985035 COLOR~            8 Tornado
## 3 19920624 1943   19920624 1943 NA           9985036 COLOR~            8 Tornado
## 4 19920625 1950   19920625 1950 NA           9985037 COLOR~            8 Thunderst~
## 5 19920626 1251   19920626 1251 NA           9985038 COLOR~            8 Tornado
## 6 19920626 1840   19920626 1840 NA           9985040 COLOR~            8 Tornado
## # ... with 7 more variables: CZ_NAME <chr>, CZ_FIPS <dbl>, SOURCE <lgl>,
## #   BEGIN_LAT <dbl>, BEGIN_LON <dbl>, END_LAT <dbl>, END_LON <dbl>
```

#Pad the state and county FIPS with a "0" at the beginning (hint: there's a function in stringr to do this) and then unite the two columns to make one fips column with the 5-digit county FIPS code (5 points)

```
STATE_FIPS <- str_pad(StormDetails_CType$STATE_FIPS, width= 3,side ="left", pad ="0")
CZ_FIPS <- str_pad(StormDetails_CType$CZ_FIPS, width= 3, side ="left",pad ="0")
newdata5 <- unite(StormDetails_CType, "FIPS","STATE_FIPS" , "CZ_FIPS", sep="")
head(newdata5)
```

```
## # A tibble: 6 x 13
##   BEGIN_DATE_TIME END_DATE_TIME EPISODE_ID EVENT_ID STATE   FIPS  EVENT_TYPE
##   <chr>           <chr>         <lgl>         <dbl> <chr>   <chr> <chr>
## 1 19920624 1511   19920624 1511 NA          9985034 COLORA~ 869   Thunderstorm ~
## 2 19920624 1827   19920624 1827 NA          9985035 COLORA~ 841   Tornado
## 3 19920624 1943   19920624 1943 NA          9985036 COLORA~ 863   Tornado
## 4 19920625 1950   19920625 1950 NA          9985037 COLORA~ 885   Thunderstorm ~
## 5 19920626 1251   19920626 1251 NA          9985038 COLORA~ 841   Tornado
## 6 19920626 1840   19920626 1840 NA          9985040 COLORA~ 8123  Tornado
## # ... with 6 more variables: CZ_NAME <chr>, SOURCE <lgl>, BEGIN_LAT <dbl>,
## #   BEGIN_LON <dbl>, END_LAT <dbl>, END_LON <dbl>
```

#Change all the column names to lower case (you may want to try the rename_all function for this) (5 points)

```
rename_all(newdata5, tolower)
```

```
## # A tibble: 13,534 x 13
##    begin_date_time end_date_time episode_id event_id state   fips  event_type
##    <chr>           <chr>         <lgl>         <dbl> <chr>   <chr> <chr>
##  1 19920624 1511   19920624 1511 NA          9985034 COLORA~ 869   Thunderstorm~
##  2 19920624 1827   19920624 1827 NA          9985035 COLORA~ 841   Tornado
##  3 19920624 1943   19920624 1943 NA          9985036 COLORA~ 863   Tornado
##  4 19920625 1950   19920625 1950 NA          9985037 COLORA~ 885   Thunderstorm~
##  5 19920626 1251   19920626 1251 NA          9985038 COLORA~ 841   Tornado
##  6 19920626 1840   19920626 1840 NA          9985040 COLORA~ 8123  Tornado
##  7 19920626 1840   19920626 1840 NA          9985041 COLORA~ 8123  Tornado
##  8 19920626 1843   19920626 1843 NA          9985042 COLORA~ 887   Tornado
##  9 19920626 2010   19920626 2010 NA          9985043 COLORA~ 875   Tornado
## 10 19920627 1428   19920627 1428 NA          9985044 COLORA~ 8121  Tornado
## # ... with 13,524 more rows, and 6 more variables: cz_name <chr>, source <lgl>,
## #   begin_lat <dbl>, begin_lon <dbl>, end_lat <dbl>, end_lon <dbl>
```

#There is data that comes with R on U.S. states (data("state")). Use that to create a dataframe with the state name, area, and region

```
data("state")
us_state_info<-data.frame(state=state.name, region=state.region, area=state.area)
us_state_info$state_upper_case=str_to_upper(us_state_info$state)
us_state_info$state <- NULL
us_state_info <- rename(.data=us_state_info, state= state_upper_case)
us_state_info
```

```
##            region   area          state
```

```
## 1           South  51609       ALABAMA
## 2            West 589757        ALASKA
## 3            West 113909       ARIZONA
## 4           South  53104      ARKANSAS
## 5            West 158693    CALIFORNIA
## 6            West 104247      COLORADO
## 7       Northeast   5009   CONNECTICUT
## 8           South   2057      DELAWARE
## 9           South  58560       FLORIDA
## 10          South  58876       GEORGIA
## 11           West   6450        HAWAII
## 12           West  83557         IDAHO
## 13 North Central  56400      ILLINOIS
## 14 North Central  36291       INDIANA
## 15 North Central  56290          IOWA
## 16 North Central  82264        KANSAS
## 17          South  40395      KENTUCKY
## 18          South  48523     LOUISIANA
## 19      Northeast  33215         MAINE
## 20          South  10577      MARYLAND
## 21      Northeast   8257 MASSACHUSETTS
## 22 North Central  58216      MICHIGAN
## 23 North Central  84068     MINNESOTA
## 24          South  47716   MISSISSIPPI
## 25 North Central  69686      MISSOURI
## 26           West 147138       MONTANA
## 27 North Central  77227      NEBRASKA
## 28           West 110540        NEVADA
## 29      Northeast   9304 NEW HAMPSHIRE
## 30      Northeast   7836    NEW JERSEY
## 31           West 121666    NEW MEXICO
## 32      Northeast  49576      NEW YORK
## 33          South  52586 NORTH CAROLINA
## 34 North Central  70665  NORTH DAKOTA
## 35 North Central  41222          OHIO
## 36          South  69919      OKLAHOMA
## 37           West  96981        OREGON
## 38      Northeast  45333  PENNSYLVANIA
## 39      Northeast   1214  RHODE ISLAND
## 40          South  31055 SOUTH CAROLINA
## 41 North Central  77047  SOUTH DAKOTA
## 42          South  42244     TENNESSEE
## 43          South 267339         TEXAS
## 44           West  84916          UTAH
## 45      Northeast   9609       VERMONT
## 46          South  40815      VIRGINIA
## 47           West  68192    WASHINGTON
## 48          South  24181 WEST VIRGINIA
## 49 North Central  56154     WISCONSIN
## 50           West  97914       WYOMING
```

#Create a dataframe with the number of events per state in the year of your birth. Merge in the state information dataframe you just created. Remove any states that are not in the state information dataframe. (5 points) #number of events per state in the year of your birth

```
newset<- data.frame(table(Stormdetails$STATE))
newset1<-rename(newset, c("state"="Var1", "num_events"="Freq"))
newset1
```

```
##                      state num_events
## 1               ALABAMA        347
## 2               ARIZONA         82
## 3              ARKANSAS        567
## 4            CALIFORNIA         33
## 5              COLORADO        266
## 6           CONNECTICUT         46
## 7              DELAWARE         23
## 8   DISTRICT OF COLUMBIA          9
## 9               FLORIDA        368
## 10              GEORGIA        393
## 11                IDAHO         68
## 12             ILLINOIS        397
## 13              INDIANA        336
## 14                 IOWA        287
## 15               KANSAS        924
## 16             KENTUCKY        156
## 17            LOUISIANA        622
## 18                MAINE         38
## 19             MARYLAND        106
## 20        MASSACHUSETTS         47
## 21             MICHIGAN        226
## 22            MINNESOTA        203
## 23          MISSISSIPPI        372
## 24             MISSOURI        414
## 25              MONTANA         70
## 26             NEBRASKA        468
## 27               NEVADA         18
## 28        NEW HAMPSHIRE          8
## 29           NEW JERSEY        106
## 30           NEW MEXICO        154
## 31             NEW YORK        268
## 32       NORTH CAROLINA        317
## 33         NORTH DAKOTA         91
## 34                 OHIO        433
## 35             OKLAHOMA       1499
## 36               OREGON         16
## 37         PENNSYLVANIA        372
## 38         RHODE ISLAND          4
## 39       SOUTH CAROLINA        186
## 40         SOUTH DAKOTA        283
## 41            TENNESSEE        186
## 42                TEXAS       2087
## 43                 UTAH         47
## 44              VERMONT         23
## 45             VIRGINIA        185
## 46           WASHINGTON         12
## 47        WEST VIRGINIA        110
## 48            WISCONSIN        187
```

```
## 49                    WYOMING          74
```

#Merge in the state information dataframe

```
merged <- merge(x=newset1,y=us_state_info,by.x="state", by.y="state")
head(merged)
```

```
##          state num_events    region   area
## 1      ALABAMA        347     South  51609
## 2      ARIZONA         82      West 113909
## 3     ARKANSAS        567     South  53104
## 4   CALIFORNIA         33      West 158693
## 5     COLORADO        266      West 104247
## 6  CONNECTICUT         46 Northeast   5009
```

#PLOT

```
library(ggplot2)
```

```
storm_plot <- ggplot(merged, aes(x=area,y=num_events)) + geom_point(aes(col = region))+ labs(x="Land are
storm_plot
```