Assignment 3

Alex Zhang

12/8/2021

R Markdown

1. read the data file

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
storm <- read_csv("C:/Users/alex_zhang4/Downloads/McDaniel College/ANA 515/StormEvents_details-ftp_v1.0_
```

```
##
## -- 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.
```

2. select variables given in assignment question 2

```
myvars <- c("BEGIN_DATE_TIME","END_DATE_TIME",
            "EPISODE_ID","EVENT_ID",
            "STATE_FIPS","STATE","CZ_NAME",
            "CZ_FIPS","CZ_TYPE",
            "EVENT_TYPE","SOURCE","BEGIN_LAT",
  "BEGIN_LON",
  "END_LAT",
  "END_LON")

newstorm <- storm[myvars]
```

3. change dates to a "date-time" class

```
library(lubridate)
```

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

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

```
newstorm %>%
  mutate(BEGIN_DATE_TIME = dmy_hms(BEGIN_DATE_TIME), END_DATE_TIME = dmy_hms(END_DATE_TIME))
```

```
## # A tibble: 7,335 x 15
##    BEGIN_DATE_TIME     END_DATE_TIME       EPISODE_ID EVENT_ID STATE_FIPS STATE
##    <dttm>              <dttm>              <lgl>         <dbl>      <dbl> <chr>
##  1 1984-02-26 23:45:00 1984-02-26 23:45:00 NA          9976714          1 ALABA~
##  2 1984-04-25 19:20:00 1984-04-25 19:20:00 NA         10067332         31 NEBRA~
##  3 1984-04-29 20:00:00 1984-04-29 20:00:00 NA         10057467         29 MISSO~
##  4 1984-08-09 14:35:00 1984-08-09 14:35:00 NA         10090123         37 NORTH~
##  5 1984-04-19 00:45:00 1984-04-19 00:45:00 NA         10046342         22 LOUIS~
##  6 1984-04-19 16:53:00 1984-04-19 16:53:00 NA          9976772          1 ALABA~
```

```
##  7 1984-04-19 16:57:00 1984-04-19 16:57:00 NA          9976773          1 ALABA~
##  8 1984-04-19 18:15:00 1984-04-19 18:15:00 NA          9976774          1 ALABA~
##  9 1984-04-19 18:15:00 1984-04-19 18:15:00 NA          9976775          1 ALABA~
## 10 1984-04-19 19:05:00 1984-04-19 19:05:00 NA          9976776          1 ALABA~
## # ... with 7,325 more rows, and 9 more variables: CZ_NAME <chr>, CZ_FIPS <dbl>,
## #   CZ_TYPE <chr>, EVENT_TYPE <chr>, SOURCE <lgl>, BEGIN_LAT <dbl>,
## #   BEGIN_LON <dbl>, END_LAT <dbl>, END_LON <dbl>
```

4. change state and county names to title case

```
newstorm$STATE = str_to_title(newstorm$STATE)
newstorm$CZ_NAME = str_to_title(newstorm$CZ_NAME)
```

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

```
filter(newstorm, CZ_TYPE == "C")
```

```
## # A tibble: 7,335 x 15
##    BEGIN_DATE_TIME    END_DATE_TIME EPISODE_ID EVENT_ID STATE_FIPS STATE CZ_NAME
##    <chr>              <chr>         <lgl>         <dbl>      <dbl> <chr> <chr>
##  1 26-FEB-84 23:45:00 26-FEB-84 23~ NA          9976714          1 Alab~ Baldwin
##  2 25-APR-84 19:20:00 25-APR-84 19~ NA         10067332         31 Nebr~ Hall
##  3 29-APR-84 20:00:00 29-APR-84 20~ NA         10057467         29 Miss~ New Ma~
##  4 09-AUG-84 14:35:00 09-AUG-84 14~ NA         10090123         37 Nort~ Pitt
##  5 19-APR-84 00:45:00 19-APR-84 00~ NA         10046342         22 Loui~ Bossier
##  6 19-APR-84 16:53:00 19-APR-84 16~ NA          9976772          1 Alab~ Bibb
##  7 19-APR-84 16:57:00 19-APR-84 16~ NA          9976773          1 Alab~ Bibb
##  8 19-APR-84 18:15:00 19-APR-84 18~ NA          9976774          1 Alab~ Elmore
##  9 19-APR-84 18:15:00 19-APR-84 18~ NA          9976775          1 Alab~ Elmore
## 10 19-APR-84 19:05:00 19-APR-84 19~ NA          9976776          1 Alab~ Tallap~
## # ... with 7,325 more rows, and 8 more variables: CZ_FIPS <dbl>, CZ_TYPE <chr>,
## #   EVENT_TYPE <chr>, SOURCE <lgl>, BEGIN_LAT <dbl>, BEGIN_LON <dbl>,
## #   END_LAT <dbl>, END_LON <dbl>
```

```
newstorm <- select(newstorm, -CZ_TYPE)
```

6. 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 or 6-digit county FIPS code

```
newstorm$STATE_FIPS <- str_pad(newstorm$STATE_FIPS,width = 3, side = "left",pad="0")
newstorm$CZ_FIPS <- str_pad(newstorm$CZ_FIPS,width = 3, side = "left",pad="0")
newstorm <- unite(newstorm,"new_FIPS",STATE_FIPS,CZ_FIPS,remove = FALSE)
```

7. Change all the column names to lower case (you may want to try the rename_all function for this)

```
newstorm <- rename_all(newstorm,tolower)
```

8. There is data that comes with base R on U.S. states (data("state")). Use that to create a dataframe with these three columns: state name, area, and region

```
state <- data("state")

us_state_info <- data.frame(state=state.name, region=state.region, area=state.area)
```

9. 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 in step 8. Remove any states that are not in the state information dataframe.

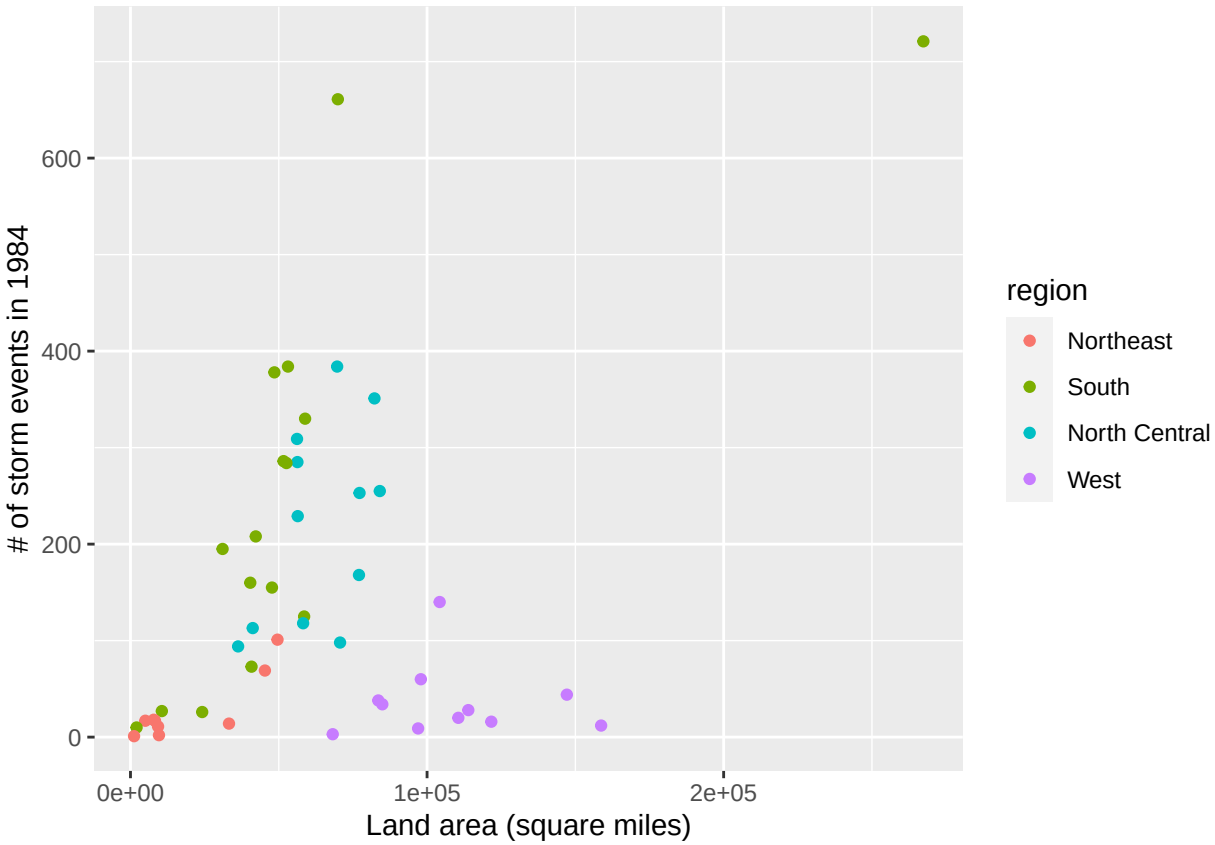
```
Newset<- data.frame(table(newstorm$state))
Newset <- rename(Newset, c("state"="Var1"))
merged <- merge(x=Newset, y=us_state_info, by.x="state",by.y="state")
view(merged)
head(merged)
```

```
##         state Freq    region   area
## 1     Alabama  286     South  51609
## 2     Arizona   28      West 113909
## 3    Arkansas  384     South  53104
## 4  California   12      West 158693
## 5    Colorado  140      West 104247
## 6 Connecticut   17 Northeast   5009
```

10. Create the following plot

```
library(ggplot2)
storm_plot <- ggplot(merged, aes(x=area,y=Freq)) +
  geom_point(aes(color = region))+
  labs(x="Land area (square miles)",
       y = "# of storm events in 1984")

storm_plot
```


600
400
200
0
# of storm events in 1984
0e+00
1e+05
2e+05
Land area (square miles)
region
Northeast
South
North Central
West