Lab 4 assessment

Tom Palmer

21st October 2017

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
# library(dplyr)
# library(ggplot2)
library(lubridate)
```

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

```
## The following object is masked from 'package:base':
##
##     date
```

```
# 1.
setwd("H:\\all\\teaching\\Math550 Stats in Practice\\R\\MATH550\\R_course\\Lab4\\Coursework")
```

```
# 2. Load
```

```
storm <- read.csv("Australia_severe_storms_1975-2015.csv",
                  stringsAsFactors=FALSE)

print("Dimensions of StormEvents data set are:")
```

```
## [1] "Dimensions of StormEvents data set are:"
```

```
print(dim(storm))
```

```
## [1] 14457    14
```

```
# 3. Clean
```

```
storm <- select(storm, Event.ID, Database, Date.Time,
                Nearest.town, State, Latitude, Longitude)


#storm$Date.Time<-dmy_hm(storm$Date.Time)


storm <- storm[storm$Database != "Waterspout", ]
print(dim(storm))
```

```
## [1] 14417     7
```

```
print(head(storm))
```

```
##   Event.ID Database        Date.Time   Nearest.town State Latitude
## 1    20812     Wind 23/11/1975 07:00         SYDNEY   NSW -33.8834
## 2    20813  Tornado 02/12/1975 14:00         BARHAM   NSW -35.6333
## 3    20814     Wind 09/01/1976 08:50 COFF'S HARBOUR   NSW -30.3167
## 4    20815     Hail 16/02/1976 14:00      BANKSTOWN   NSW -33.8834
## 5    20816     Rain 25/10/1976 14:00          BOOMI   NSW -28.4333
## 6    20817     Hail 08/11/1976 14:00          YOUNG   NSW -34.3167
##   Longitude
## 1  151.2167
## 2  144.1333
## 3  153.1167
## 4  151.2167
## 5  152.6167
## 6  148.3000
```

```
# 4. time zone allocation
```

```
# OlsonNames() classifications

state_allocate <- function(state.in="NSW",town.in="Sydney")
{
  if(state.in=="QLD") timezone="Australia/Queensland"
  if(state.in=="VIC") timezone="Australia/Victoria"
  if(state.in=="SA") timezone="Australia/South"
  if(state.in=="WA") timezone="Australia/West"
  if(state.in=="TAS") timezone="Australia/Tasmania"
  if(state.in=="NT") timezone="Australia/North"
  if(state.in=="ACT") timezone="Australia/ACT"
  if(state.in=="NSW") timezone="Australia/NSW"
  if((state.in=="NSW") & ((town.in=="Broken Hill") | (town.in=="BROKEN HILL") |
                       (town.in=="BROKEN HILL AREA") |
                       (town.in=="MILDURA/BROKEN HILL") |
                       (town.in=="BROKEN HILL AIRPORT"))) timezone="Australia/Broken_Hill"
 return(timezone)
}

# Well done to those who allocated all Broken Hill categories
# Marks awarded for Broken Hill or BROKEN HILL
#
#
# Using strings which we learn about in lab 5 there are
# easier ways to extract all occurrences


storm$time.zone <- character(14417)
for(i in 1:14417) {
  storm$time.zone[i] <- state_allocate(storm$State[i], storm$Nearest.town[i])
}

table(storm$time.zone)
```

```
##
```

```
## Australia/Broken_Hill       Australia/North         Australia/NSW
##                    48                   535                  4467
##  Australia/Queensland       Australia/South    Australia/Tasmania
##                  2760                  2090                   201
##    Australia/Victoria        Australia/West
##                  2250                  2066
```

```
# 5. UTC time
```

```
for(i in 1:nrow(storm)){
  storm$UTC.Time[i] <- with_tz(dmy_hm(storm$Date.Time[i], tz=storm$time.zone[i]), "UTC")
}
print(head(storm))
```

```
##   Event.ID Database        Date.Time   Nearest.town State Latitude
## 1    20812     Wind 23/11/1975 07:00         SYDNEY   NSW -33.8834
## 2    20813  Tornado 02/12/1975 14:00         BARHAM   NSW -35.6333
## 3    20814     Wind 09/01/1976 08:50 COFF'S HARBOUR   NSW -30.3167
## 4    20815     Hail 16/02/1976 14:00      BANKSTOWN   NSW -33.8834
## 5    20816     Rain 25/10/1976 14:00          BOOMI   NSW -28.4333
## 6    20817     Hail 08/11/1976 14:00          YOUNG   NSW -34.3167
##   Longitude     time.zone  UTC.Time
## 1  151.2167 Australia/NSW 185918400
## 2  144.1333 Australia/NSW 186721200
## 3  153.1167 Australia/NSW 189985800
## 4  151.2167 Australia/NSW 193287600
## 5  152.6167 Australia/NSW 215064000
## 6  148.3000 Australia/NSW 216270000
```

```
#
# Alternative approach:
# Some people used the following approach using vectorisation
# which is particularly fast.
#

storm$UTC.Time <- dmy_hm("01/01/1970 00:00", tz="UTC")

for(timezone in unique(storm$time.zone)){
  storm[storm$time.zone == timezone, "UTC.Time"] <-
    dmy_hm(storm[storm$time.zone == timezone, "Date.Time"], tz=timezone)
}
print(head(storm))
```

```
##   Event.ID Database        Date.Time   Nearest.town State Latitude
## 1    20812     Wind 23/11/1975 07:00         SYDNEY   NSW -33.8834
## 2    20813  Tornado 02/12/1975 14:00         BARHAM   NSW -35.6333
## 3    20814     Wind 09/01/1976 08:50 COFF'S HARBOUR   NSW -30.3167
## 4    20815     Hail 16/02/1976 14:00      BANKSTOWN   NSW -33.8834
## 5    20816     Rain 25/10/1976 14:00          BOOMI   NSW -28.4333
## 6    20817     Hail 08/11/1976 14:00          YOUNG   NSW -34.3167
##   Longitude     time.zone            UTC.Time
## 1  151.2167 Australia/NSW 1975-11-22 20:00:00
## 2  144.1333 Australia/NSW 1975-12-02 03:00:00
## 3  153.1167 Australia/NSW 1976-01-08 21:50:00
## 4  151.2167 Australia/NSW 1976-02-16 03:00:00
## 5  152.6167 Australia/NSW 1976-10-25 04:00:00
```

```
## 6  148.3000 Australia/NSW 1976-11-08 03:00:00
```

```
# 6. Month and year
```

```
storm$month <- month(storm$UTC.Time)
storm$year <- year(storm$UTC.Time)
head(storm)
```

```
##   Event.ID Database        Date.Time   Nearest.town State Latitude
## 1    20812     Wind 23/11/1975 07:00         SYDNEY   NSW -33.8834
## 2    20813  Tornado 02/12/1975 14:00         BARHAM   NSW -35.6333
## 3    20814     Wind 09/01/1976 08:50 COFF'S HARBOUR   NSW -30.3167
## 4    20815     Hail 16/02/1976 14:00      BANKSTOWN   NSW -33.8834
## 5    20816     Rain 25/10/1976 14:00          BOOMI   NSW -28.4333
## 6    20817     Hail 08/11/1976 14:00          YOUNG   NSW -34.3167
##   Longitude     time.zone            UTC.Time month year
## 1  151.2167 Australia/NSW 1975-11-22 20:00:00    11 1975
## 2  144.1333 Australia/NSW 1975-12-02 03:00:00    12 1975
## 3  153.1167 Australia/NSW 1976-01-08 21:50:00     1 1976
## 4  151.2167 Australia/NSW 1976-02-16 03:00:00     2 1976
## 5  152.6167 Australia/NSW 1976-10-25 04:00:00    10 1976
## 6  148.3000 Australia/NSW 1976-11-08 03:00:00    11 1976
```

```
# 7. Events
```

```
table(storm$Database)
```

```
##
##     Hail Lighting     Rain  Tornado     Wind
##     3098      129     4504      676     6010
```

```
Events <- c("Hail", "Lighting", "Rain", "Tornado", "Wind")

Events_count <- data.frame(Month=seq(1,12))
Events_count
```

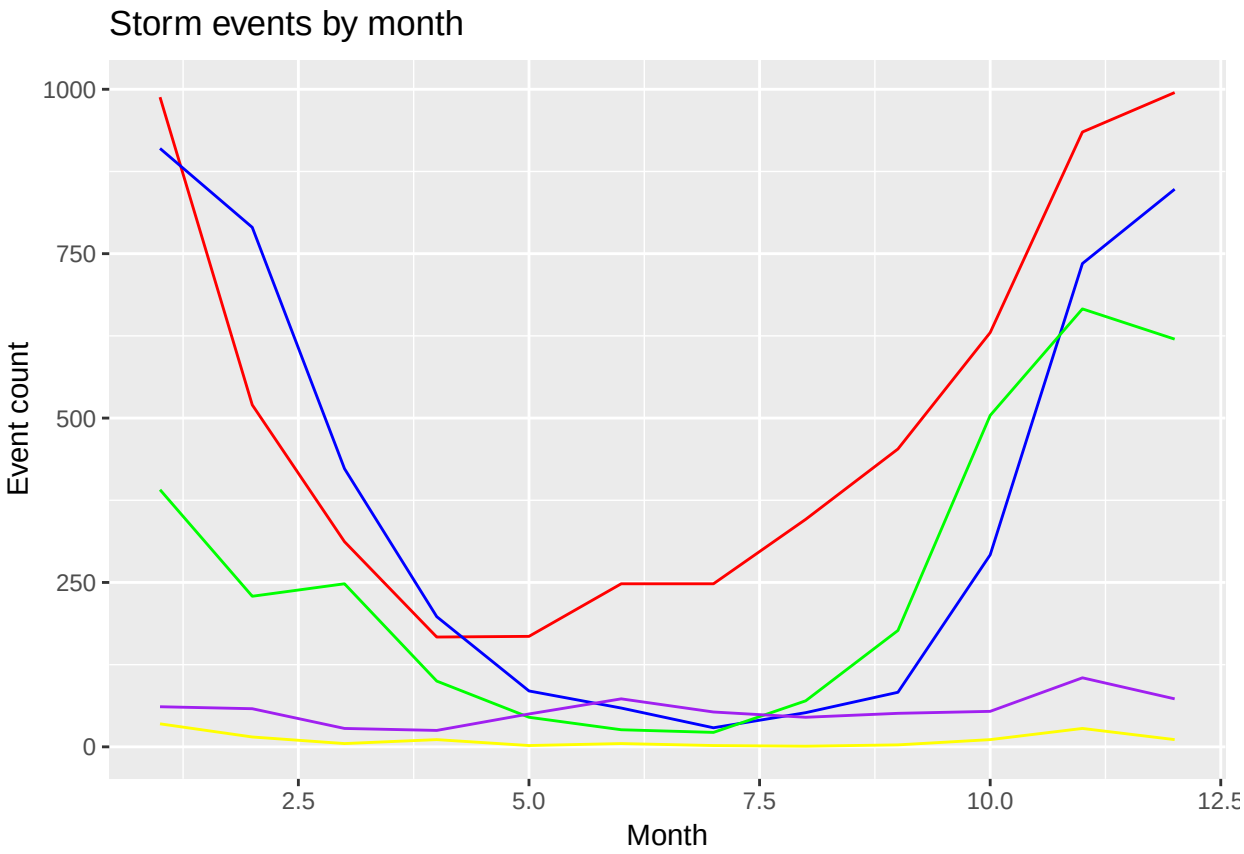
```
##    Month
## 1      1
## 2      2
## 3      3
## 4      4
## 5      5
## 6      6
## 7      7
## 8      8
## 9      9
## 10    10
## 11    11
## 12    12
```

```
Events_count$Hail <- NA
for(i in 1:12) {
  Events_count$Hail[i] <- sum(storm$month==i & storm$Database=="Hail")
}

Events_count$Lighting <- NA
for(i in 1:12) {
  Events_count$Lighting[i] <- sum(storm$month==i & storm$Database=="Lighting")
```

```
}

Events_count$Rain <- NA
for(i in 1:12) {
  Events_count$Rain[i] <- sum(storm$month==i & storm$Database=="Rain")
}

Events_count$Tornado <- NA
for(i in 1:12) {
  Events_count$Tornado[i] <- sum(storm$month==i & storm$Database=="Tornado")
}

Events_count$Wind <- NA
for(i in 1:12) {
  Events_count$Wind[i] <- sum(storm$month==i & storm$Database=="Wind")
}

Events_count
```

```
##    Month Hail Lighting Rain Tornado Wind
## 1      1  391       35  910      61  988
## 2      2  229       15  790      58  520
## 3      3  248        5  423      28  312
## 4      4  100       11  198      25  167
## 5      5   45        2   85      50  168
## 6      6   26        5   59      73  248
## 7      7   22        2   29      53  248
## 8      8   70        1   52      45  346
## 9      9  177        3   83      51  453
## 10    10  504       11  292      54  630
## 11    11  666       28  735     105  935
## 12    12  620       11  848      73  995
```

```
q7plot <- ggplot(Events_count) +
   geom_line(aes(x=Month, y=Wind), colour="red") +
   geom_line(aes(x=Month, y=Rain), colour="blue") +
   geom_line(aes(x=Month, y=Hail), colour="green") +
   geom_line(aes(x=Month, y=Tornado), colour="purple") +
   geom_line(aes(x=Month, y=Lighting), colour="yellow") +
  labs(x="Month", y="Event count", title="Storm events by month")
q7plot
```


Storm events by month
1000
750
500
250
0
Event count
2.5
5.0
7.5
10.0
12.5
Month