Project 2 Solutions

(Abhimanyu Agarwal)

Collaborators: N/A

TA help:

1) Hilda Ibriga - Question 4 and 5, Helped me understanding and applying concepts in R.

2) Evan Li, Cross-checked my knitted pdf to enhance result reporting professionally.

Online resources used: N/A

Question 1

```
#Loads into dataframe called myDF
myDF<-read.csv("/class/datamine/data/disney/metadata.csv")

#Prints the first 1 line using head() just to see
head(myDF, n = 1)
```

```
        DATE WDW_TICKET_SEASON DAYOFWEEK DAYOFYEAR WEEKOFYEAR MONTHOFYEAR
1 01/01/2015                           5         0          0           1
  YEAR         SEASON HOLIDAYPX HOLIDAYM HOLIDAYN HOLIDAY WDWTICKETSEASON
1 2015 CHRISTMAS PEAK         0        5      nyd       1
  WDWRaceN WDWeventN WDWevent WDWrace      WDWSEASON WDWMAXTEMP WDWMINTEMP
1                          0       0 CHRISTMAS PEAK      73.02      59.81
  WDWMEANTEMP MKeventN MKevent EPeventN EPevent HSeventN HSevent AKeventN
1       66.41                0                0                0
  AKevent HOLIDAYJ inSession inSession_Enrollment inSession_wdw
1       0                 0%                   0%            0%
  inSession_dlr inSession_sqrt_WDW inSession_sqrt_DLR inSession_California
1            0%                 0%                 0%                    0%
  inSession_DC inSession_Central_FL inSession_Drive1_FL
1           0%                   0%                  0%
  inSession_Drive2_FL inSession_Drive_CA inSession_Florida
1                  0%                 0%                0%
  inSession_Mardi_Gras inSession_Midwest inSession_NY_NJ
1                   0%                0%              0%
  inSession_NY_NJ_PA inSession_New_England inSession_New_Jersey
1                 0%                    0%                   0%
  inSession_Nothwest INSESSION_PLANES inSession_SoCal inSession_Southwest
1                 0%                0%             0%                   0%
  SUNSET_WDW MKEMHMORN MKEMHMYEST MKEMHMTOM MKEMHEVE MKHOURSEMH
1      17:42         1          1         0        0         18
  MKHOURSEMHYEST MKHOURSEMHTOM MKEMHEYEST MKEMHETOM EPEMHMORN EPEMHMYEST
1             19            17          0         0         0          0
```

```
  EPEMHMTOM EPEMHEVE EPEMHEYEST EPEMHETOM EPHOURSEMH EPHOURSEMHYEST
1         0        0          0         1         13             17
  EPHOURSEMHTOM HSEMHMORN HSEMHMYEST HSEMHMTOM HSEMHEVE HSEMHEYEST
1            15         0          0         0        0          1
  HSEMHETOM HSHOURSEMH HSHOURSEMHYEST HSHOURSEMHTOM AKEMHMORN AKEMHMYEST
1         0         14             16            14         0          1
  AKEMHMTOM AKEMHEVE AKEMHEYEST AKEMHETOM AKHOURSEMH AKHOURSEMHYEST
1         0        0          0         0         11             15
  AKHOURSEMHTOM MKOPEN MKCLOSE MKHOURS MKEMHOPEN MKEMHCLOSE MKOPENYEST
1            12   8:00   25:00      17      7:00      25:00       8:00
  MKCLOSEYEST MKHOURSYEST MKOPENTOM MKCLOSETOM MKHOURSTOM EPOPEN EPCLOSE
1       26:00          18      8:00      25:00         17   8:00   21:00
  EPHOURS EPEMHOPEN EPEMHCLOSE EPOPENYEST EPCLOSEYEST EPHOURSYEST
1      13      8:00      21:00       8:00       25:00          17
  EPOPENTOM EPCLOSETOM EPHOURSTOM HSOPEN HSCLOSE HSHOURS HSEMHOPEN
1      8:00      21:00         13   8:00   22:00      14      8:00
  HSEMHCLOSE HSOPENYEST HSCLOSEYEST HSHOURSYEST HSOPENTOM HSCLOSETOM
1      22:00       8:00       24:00          16      8:00      22:00
  HSHOURSTOM AKOPEN AKCLOSE AKHOURS AKEMHOPEN AKEMHCLOSE AKOPENYEST
1         14   8:00   19:00      11      8:00      19:00       8:00
  AKCLOSEYEST AKHOURSYEST AKOPENTOM AKCLOSETOM AKHOURSTOM
1       22:00          14      8:00      20:00         12
  MKDAYSBEFORENONEVENT MKDAYSSINCENONEVENT MKEVENTSTREAK MKEVENTSTREAK_F
1                    1                   1             0               0
  PARTYSEASON_WDW WDWMINTEMP_mean WEATHER_WDWHIGH WEATHER_WDWLOW
1            NONE        53.37571            70.3           50.2
  WEATHER_WDWPRECIP CapacityLost_MK CapacityLost_EP CapacityLost_HS
1              0.12          616246          367265          296273
  CapacityLost_AK CapacityLostWGT_MK CapacityLostWGT_EP CapacityLostWGT_HS
1          236654           53904354           34718635           26907827
  CapacityLostWGT_AK EP09CAPACITY HS20CAPACITY MKPRDDAY MKPRDDT1 MKPRDDT2
1           20971646         1600         1000        2    12:00    15:30
                           MKPRDDN MKPRDNGT MKPRDNT1 MKPRDNT2
1 Disney Festival of Fantasy Parade        1    22:15
                         MKPRDNN MKFIREWK MKFIRET1 MKFIRET2
1 Main Street Electrical Parade        1    21:00
                        MKFIREN EPFIREWK EPFIRET1 EPFIRET2
1 Wishes Nighttime Spectacular        1    21:00
                              EPFIREN HSPRDDAY HSPRDDT1 HSPRDDN HSFIREWK
1 IllumiNations: Reflections of Earth        0       NA      NA        0
  HSFIRET1 HSFIRET2 HSFIREN HSSHWNGT HSSHWNT1 HSSHWNT2    HSSHWNN
1                                3    18:30    20:00 Fantasmic!
  HSFIREWKS AKPRDDAY AKPRDDT1 AKPRDDT2 AKPRDDN AKFIREN AKSHWNGT AKSHWNT1
1         1        0       NA       NA      NA      NA        0
  AKSHWNT2 AKSHWNN
1
```

#Note: TA suggested to print a few lines just because there are so many columns to print otherwise.

Question 2

```
#Given code in project
our_vec <- myDF$WDWMAXTEMP
```

```r
#Prints the first value
firstval <- our_vec[1]
head(firstval)
```

```
[1] 73.02
```

```r
#Printed value is: 73.02

#Prints the 50th value
fiftyval <- our_vec[50]
head(fiftyval)
```

```
[1] 51.24
```

```r
#Printed value is: 51.24

#Type of a vector
typeof(our_vec)
```

```
[1] "double"
```

```r
#Type is "double"
```

Question 3

```r
#First 50 values
first50 <- head(our_vec, n = 50)
head(first50)
```

```
[1] 73.02 78.00 83.12 83.93 72.30 77.67
```

```r
#Prints values of first 50 elements

#Last 50 values
last50 <- tail(our_vec, n = 50)
head(last50)
```

```
[1] 78.73 78.38 75.73 74.79 76.84 81.25
```

```r
#Prints the values of last 50 elements

#Adding the two element vectors
mix <- first50 + last50
head(mix)
```

```
[1] 151.75 156.38 158.85 158.72 149.14 158.92
```

```r
#Prints the values of after performing addition
```

Question 4

```r
#Stores the temperature values greater or equal to 80. Since our_vec is already defined
hot <- our_vec[our_vec >= 80]
head(hot)
```

```
[1] 83.12 83.93 80.63 81.34 81.95 82.33
```

```r
#Prints some values
```

```r
#Tells us how many elements are in hot (vector)
length(hot)
```

```
[1] 1255
```

```r
#Sum of hot and first50
result <- hot + first50
```

```
Warning in hot + first50: longer object length is not a multiple of shorter
object length
```

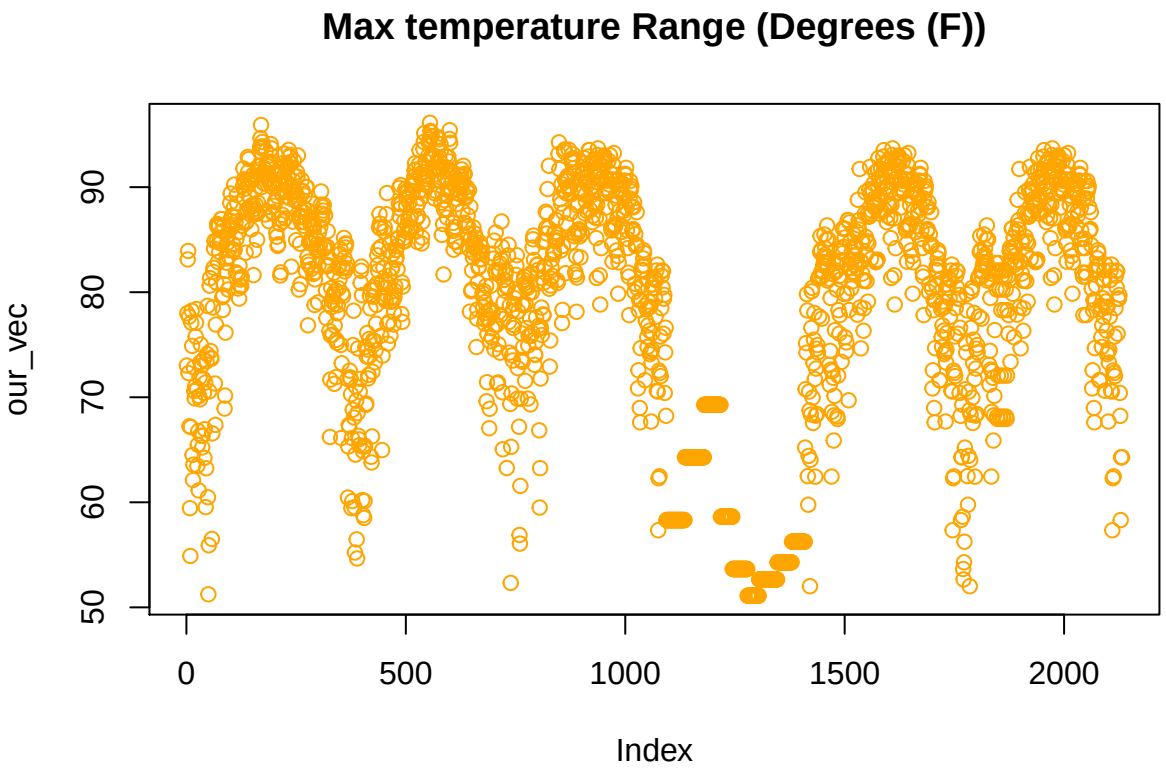
Explanantion : We get an error message that is because of the length of the two vectors is not the same. When we perform addition it adds the vectors element by element which results in a vector. Since the lengths of the two vectors here are different, we get a warning message as seen. R performs recycling on the shorter one by making the two vectors equal. This addition also results in a vector.

```r
#prints result
head(result)
```

```
[1] 156.14 161.93 163.75 165.27 154.25 160.00
```

Question 5

```r
#Plotting the values captured by the vector
plot(our_vec, main='Max temperature Range (Degrees (F))', col='orange')
```



Question 6

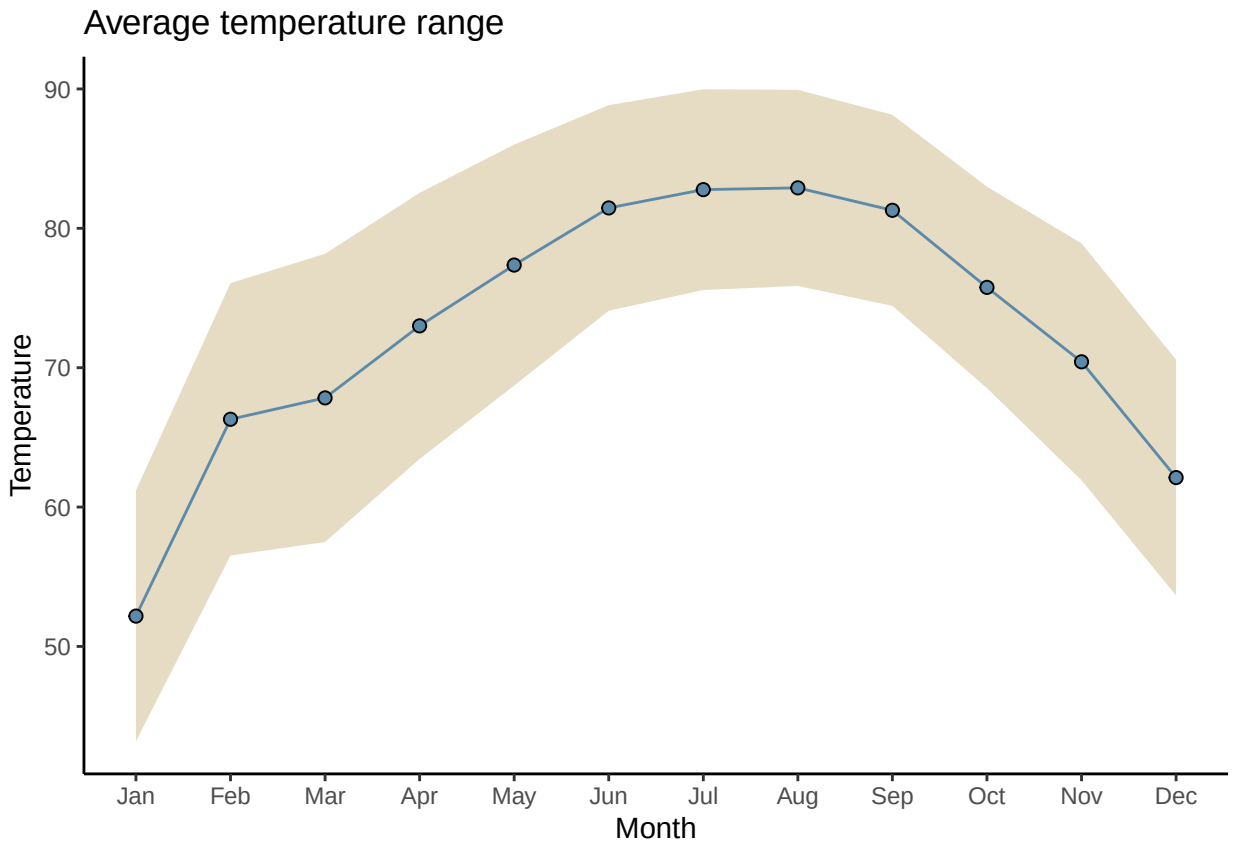
```r
#Favorite graph would be the second one. The code is below as given:
library(ggplot2)
library(tidyverse)
```

```
-- Attaching packages ------------------------------------------------- tidyverse 1.3.0 --

v tibble  2.1.3     v dplyr   1.0.2
v tidyr   1.1.1     v stringr 1.4.0
v readr   1.3.1     v forcats 0.4.0
v purrr   0.3.4

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

```r
summary_temperatures <- myDF %>%
  select(MONTHOFYEAR,WDWMAXTEMP:WDWMEANTEMP) %>%
  group_by(MONTHOFYEAR) %>%
  summarise_all(mean, na.rm=T)
ggplot(summary_temperatures, aes(x=MONTHOFYEAR)) +
  geom_ribbon(aes(ymin = WDWMINTEMP, ymax = WDWMAXTEMP), fill = "#ceb888", alpha=.5) +
  geom_line(aes(y = WDWMEANTEMP), col="#5D8AA8") +
  geom_point(aes(y = WDWMEANTEMP), pch=21,fill = "#5D8AA8", size=2) +
  theme_classic() +
  labs(x = 'Month', y = 'Temperature', title = 'Average temperature range' ) +
  scale_x_continuous(breaks=1:12, labels=month.abb)
```



1. It is my favorite graphic because it is extremely detailed making it reader friendly open to interpretation and does not show absurd variations. It discusses the topic of Average Temperature Range. Moreover, this plot does not convey all information properly.
2. It could be improvised by adding a legend, smoothing the variation could give a better sense of the Temperature Vs Month variation.
3. A fascinating aspect that attracted my attention was the bell shape of the plot with the peak being at

approximately 80 Degrees(F).

Submitting deliverables: project02.RMD, project02.R and project02.pdf

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - We are Purdue.