Project 3 Solutions

Kevin Choe

Collaborators: N/A

TA help: Hilda Ibriga, guided with problems 1-6, explained new concepts

Online resources used: Stat 190 Example Book (All problems)

Question 1

```
splash_mountain <- read.csv("/class/datamine/data/disney/splash_mountain.csv")
str (splash_mountain)
```

```
'data.frame':   223936 obs. of  4 variables:
 $ date    : chr  "01/01/2015" "01/01/2015" "01/01/2015" "01/01/2015" ...
 $ datetime: chr  "2015-01-01 07:51:12" "2015-01-01 08:02:13" "2015-01-01 08:09:12" "2015-01-01 08:16:1
 $ SACTMIN : int  NA NA NA NA NA NA NA NA NA 4 ...
 $ SPOSTMIN: int  5 5 5 5 5 5 5 5 5 NA ...
```

```
dim (splash_mountain)
```

```
[1] 223936      4
```

There are 223,936 rows and and 4 columns in the dataset

Question 2

```
mean (splash_mountain$SPOSTMIN,na.rm=T)
```

```
[1] -71.70373
```

```
sqrt (var (splash_mountain$SPOSTMIN,na.rm=T))
```

```
[1] 328.0586
```

The results of running the R code is -71.70373 for the mean and 328.0586 for the standard deviation. The results do not make sense because time is negative. I am getting a negative answer because the value is indicated that the ride as being closed.

Question 3

```
mean (splash_mountain$SPOSTMIN,na.rm=T)
```

```
[1] -71.70373
```

```
mean(splash_mountain$SPOSTMIN [which (splash_mountain$SPOSTMIN>-999)],na.rm=T)
```

```
[1] 43.3892
```

```
sqrt (var (splash_mountain$SPOSTMIN,na.rm=T))
```

```
[1] 328.0586
```

```
sqrt (var (splash_mountain$SPOSTMIN [which (splash_mountain$SPOSTMIN>-999)],na.rm=T))
```

```
[1] 31.74894
```

The results of running the R code is 43.3892 for the mean and 31.74894 for the standard deviation. The value looks reasonable now and seems to have fixed our problem.

Question 4

```
names(splash_mountain)
```

```
[1] "date"     "datetime" "SACTMIN"  "SPOSTMIN"
```

```
which(names(splash_mountain) == "SACTMIN")
```

```
[1] 3
```

```
names(splash_mountain)[3] <- "actual_wait_time"

names(splash_mountain)
```

```
[1] "date"             "datetime"         "actual_wait_time"
[4] "SPOSTMIN"
```

```
which(names(splash_mountain) == "SPOSTMIN")
```

```
[1] 4
```

```
names(splash_mountain)[4] <- "posted_min_wait_time"
```

The output from executing names(splash_mountain) is: [1] “date” “datetime” “SACTMIN” “SPOSTMIN”.

Question 5

```
?as.Date
as.Date(head(splash_mountain$date),"%m/%d/%Y")
```

```
[1] "2015-01-01" "2015-01-01" "2015-01-01" "2015-01-01" "2015-01-01"
[6] "2015-01-01"
```

```
myresults <- cut(as.Date(splash_mountain$date, "%m/%d/%Y"), "quarter")
nlevels(myresults)
```

```
[1] 20
```

```
levels(myresults) <- paste0("q",1:nlevels(myresults))
head(myresults)
```

```
[1] q1 q1 q1 q1 q1 q1
20 Levels: q1 q2 q3 q4 q5 q6 q7 q8 q9 q10 q11 q12 q13 q14 q15 q16 ... q20
```

```
splash_mountain$quarters <- myresults
head(splash_mountain)
```

```
        date            datetime actual_wait_time posted_min_wait_time
1 01/01/2015 2015-01-01 07:51:12               NA                    5
2 01/01/2015 2015-01-01 08:02:13               NA                    5
3 01/01/2015 2015-01-01 08:09:12               NA                    5
4 01/01/2015 2015-01-01 08:16:12               NA                    5
```

```
5 01/01/2015 2015-01-01 08:23:12               NA                    5
6 01/01/2015 2015-01-01 08:29:12               NA                    5
  quarters
1       q1
2       q1
3       q1
4       q1
5       q1
6       q1
```

```
tail(splash_mountain)
```

```
             date            datetime actual_wait_time
223931 12/31/2019 2020-01-01 00:27:02               NA
223932 12/31/2019 2020-01-01 00:34:02               NA
223933 12/31/2019 2020-01-01 00:41:02               NA
223934 12/31/2019 2020-01-01 00:48:02               NA
223935 12/31/2019 2020-01-01 00:55:02               NA
223936 12/31/2019 2020-01-01 01:01:02               NA
       posted_min_wait_time quarters
223931                    5      q20
223932                    5      q20
223933                    5      q20
223934                    5      q20
223935                    5      q20
223936                    5      q20
```

There are 20 quarters in the new quarter column.

Question 6

I acknowledge that the STAT 19000/29000/39000 1-credit Data Mine seminar will be recorded and posted on Piazza, for participants in this course.

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.