

Project 3 -- Seyi Ogunmodede

Instructor: Dr. Ward Daniel

- The videos help a lot

Collaboration:

Question 1

```
In [3]: options(jupyter.rich_display = F)
```

```
In [4]: list.files("/anvil/projects/tdm/data/disney")
```

```
[1] "7_dwarfs_train.csv"           "alien_saucers.csv"  
[3] "dinosaur.csv"                "entities.csv"  
[5] "expedition EVEREST.csv"     "flight_of_passage.csv"  
[7] "kilimanjaro_safaris.csv"    "metadata.csv"  
[9] "navi_river.csv"              "pirates_of_caribbean.csv"  
[11] "rock_n_rollercoaster.csv"   "slinky_dog.csv"  
[13] "soarin.csv"                 "spaceship_earth.csv"  
[15] "splash_mountain.csv"       "total.parquet"  
[17] "touringplans_data_dictionary.xlsx" "toy_story_mania.csv"  
[19] "update.py"
```

```
In [5]: myDF <- read.csv("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv",  
                      stringsAsFactors = TRUE)
```

```
In [6]: head(myDF)
```

	date	datetime	SACTMIN	SPOSTMIN
1	01/01/2015	2015-01-01 07:51:12	NA	45
2	01/01/2015	2015-01-01 08:02:13	NA	60
3	01/01/2015	2015-01-01 08:05:30	54	NA
4	01/01/2015	2015-01-01 08:09:12	NA	60
5	01/01/2015	2015-01-01 08:16:12	NA	60
6	01/01/2015	2015-01-01 08:22:16	55	NA

```
In [7]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$size
```

```
[1] 11696214
```

```
In [8]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$isdir
```

```
[1] FALSE
```

```
In [9]: file.info("/anvil/projects/tdm/data/disney/")$isdir
```

```
[1] TRUE
```

```
In [11]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$mode
```

```
[1] "664"
```

```
In [12]: file.info("/anvil/projects/tdm/data/disney/")$mode
```

```
[1] "2775"

In [13]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$mtime
[1] "2022-02-22 12:13:32 EST"

In [14]: file.info("/anvil/projects/tdm/data/disney/")$mtime
[1] "2022-02-24 17:26:22 EST"

In [15]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$ctime
[1] "2022-07-08 09:39:24 EDT"

In [16]: file.info("/anvil/projects/tdm/data/disney/")$ctime
[1] "2022-07-08 09:39:33 EDT"

In [17]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$atime
[1] "2022-09-24 21:52:46 EDT"

In [18]: file.info("/anvil/projects/tdm/data/disney/")$atime
[1] "2022-09-24 21:52:43 EDT"

In [19]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$uid
[1] 7095091

In [20]: file.info("/anvil/projects/tdm/data/disney/")$uid
[1] 7095091

In [21]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$gid
[1] 7500002

In [22]: file.info("/anvil/projects/tdm/data/disney/")$gid
[1] 7500002

In [23]: file.info("/anvil/projects/tdm/data/disney/7_dwarfs_train.csv")$uname
[1] NA

In [24]: file.info("/anvil/projects/tdm/data/disney/")$uname
[1] NA

In [25]: dim(myDF)
[1] 321631      4

In [26]: str(myDF)
'data.frame': 321631 obs. of 4 variables:
 $ date    : Factor w/ 2334 levels "01/01/2015","01/01/2016",...: 1 1 1 1 1 1 1 1 1 ...
 ... 
 $ datetime: Factor w/ 321586 levels "2015-01-01 07:51:12",...: 1 2 3 4 5 6 7 8 9 10 ...
 ... 
 $ SACTMIN : int NA NA 54 NA NA 55 NA NA NA NA ...
 $ SPOSTMIN: int 45 60 NA 60 60 NA 60 90 105 ...
```

In [27]: `head(myDF, n=21)`

	date	datetime	SACTMIN	SPOSTMIN
1	01/01/2015	2015-01-01 07:51:12	NA	45
2	01/01/2015	2015-01-01 08:02:13	NA	60
3	01/01/2015	2015-01-01 08:05:30	54	NA
4	01/01/2015	2015-01-01 08:09:12	NA	60
5	01/01/2015	2015-01-01 08:16:12	NA	60
6	01/01/2015	2015-01-01 08:22:16	55	NA
7	01/01/2015	2015-01-01 08:23:12	NA	60
8	01/01/2015	2015-01-01 08:29:12	NA	60
9	01/01/2015	2015-01-01 08:37:13	NA	90
10	01/01/2015	2015-01-01 08:44:11	NA	105
11	01/01/2015	2015-01-01 08:50:33	NA	105
12	01/01/2015	2015-01-01 08:51:12	NA	105
13	01/01/2015	2015-01-01 09:02:13	NA	105
14	01/01/2015	2015-01-01 09:09:16	NA	105
15	01/01/2015	2015-01-01 09:16:12	NA	120
16	01/01/2015	2015-01-01 09:23:12	NA	120
17	01/01/2015	2015-01-01 09:29:12	NA	120
18	01/01/2015	2015-01-01 09:34:32	NA	120
19	01/01/2015	2015-01-01 09:37:12	NA	80
20	01/01/2015	2015-01-01 09:44:11	NA	80
21	01/01/2015	2015-01-01 09:51:12	NA	80

In [28]: `tail(myDF)`

	date	datetime	SACTMIN	SPOSTMIN
321626	12/28/2021	2021-12-28 22:36:08	NA	45
321627	12/28/2021	2021-12-28 22:42:15	NA	45
321628	12/28/2021	2021-12-28 22:48:12	NA	45
321629	12/28/2021	2021-12-28 22:54:10	NA	45
321630	12/28/2021	2021-12-28 22:54:28	10	NA
321631	12/28/2021	2021-12-28 22:57:34	NA	35

In []:

In []:

There are 321631 Rows and 4 Columns. Date and datetime has Factor while SACTMIN & SPOSTMIN has integers.

Question 2

In [29]: `head(myDF$SPOSTMIN)`

```
[1] 45 60 NA 60 60 NA
```

In [30]: `head(myDF$SPOSTMIN, n=100)`

```
[1] 45 60 NA 60 60 NA 60 60 90 105 105 105 105 120 120 120 120
[19] 80 80 80 90 90 90 90 90 80 80 80 80 80 80 80 90 90
[37] 90 100 100 120 150 120 120 120 120 120 120 120 120 120 120 120 150
[55] 180 160 140 160 150 160 160 180 160 160 180 160 160 160 140 140 150 150
[73] 150 150 150 150 150 160 160 160 150 130 130 130 130 120 120 110 130
[91] 110 110 110 100 100 100 100 100 100 100 100 100 100 100 100 105
```

In [31]: `table(myDF$SPOSTMIN)`

-999	0	5	10	15	20	25	30	35	40	45	50	55
23062	32	897	1306	914	3638	2166	13939	4894	14941	15149	19555	6989
60	65	70	75	80	85	90	95	100	105	110	115	120
29034	10345	23496	12101	22971	5301	23813	6853	10773	8164	11909	3000	13097
125	130	135	140	145	150	155	160	165	170	175	180	185
1954	5982	1723	4151	940	3793	394	1739	520	1031	272	1644	163
190	195	200	205	210	215	220	225	230	235	240	250	260
375	128	249	100	257	11	88	26	26	17	37	17	10
270	280	300										
6	2	2										

In [32]: `table(myDF$SPOSTMIN, useNA="always")`

-999	0	5	10	15	20	25	30	35	40	45	50	55
23062	32	897	1306	914	3638	2166	13939	4894	14941	15149	19555	6989
60	65	70	75	80	85	90	95	100	105	110	115	120
29034	10345	23496	12101	22971	5301	23813	6853	10773	8164	11909	3000	13097
125	130	135	140	145	150	155	160	165	170	175	180	185
1954	5982	1723	4151	940	3793	394	1739	520	1031	272	1644	163
190	195	200	205	210	215	220	225	230	235	240	250	260
375	128	249	100	257	11	88	26	26	17	37	17	10
270	280	300	<NA>									
6	2	2										

In [33]: `table(myDF$SPOSTMIN)`

-999	0	5	10	15	20	25	30	35	40	45	50	55
23062	32	897	1306	914	3638	2166	13939	4894	14941	15149	19555	6989
60	65	70	75	80	85	90	95	100	105	110	115	120
29034	10345	23496	12101	22971	5301	23813	6853	10773	8164	11909	3000	13097
125	130	135	140	145	150	155	160	165	170	175	180	185
1954	5982	1723	4151	940	3793	394	1739	520	1031	272	1644	163
190	195	200	205	210	215	220	225	230	235	240	250	260
375	128	249	100	257	11	88	26	26	17	37	17	10
270	280	300										
6	2	2										

In [34]: `length(myDF$SPOSTMIN)`

[1] 321631

In [84]: `summary(myDF$SPOSTMIN)`

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
-999.000	45.000	70.000	-2.064	95.000	300.000	7635

In [85]: `head(is.na(myDF$SPOSTMIN), n=50)`

```
[1] FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE FALSE FALSE
[13] FALSE FALSE
[25] FALSE FALSE
[37] FALSE FALSE
[49] FALSE FALSE
```

In [86]: `sum(is.na(myDF$SPOSTMIN))`

```
# remember that when we sum TRUE and FALSE values
# the TRUE values become 1 and the FALSE values become 0
# so the sum is the number of TRUE values
```

```
[1] 7635
```

In [35]: *# wrong way! note.*

```
newDF <- subset(myDF, is.na(myDF$SPOSTMIN))
```

In [88]: `head(newDF)`

	date	datetime	SACTMIN	SPOSTMIN	weekday	month
1	01/01/2015	2015-01-01 07:51:12	NA	45	Thu	Jan
2	01/01/2015	2015-01-01 08:02:13	NA	60	Thu	Jan
4	01/01/2015	2015-01-01 08:09:12	NA	60	Thu	Jan
5	01/01/2015	2015-01-01 08:16:12	NA	60	Thu	Jan
7	01/01/2015	2015-01-01 08:23:12	NA	60	Thu	Jan
8	01/01/2015	2015-01-01 08:29:12	NA	60	Thu	Jan

In [89]: *# right way!*

put exclamation mark, it will switch things around.

```
newDF <- subset(myDF, !is.na(myDF$SPOSTMIN))
```

In [78]: `head(newDF)`

	date	datetime	SACTMIN	SPOSTMIN	weekday	month
1	01/01/2015	2015-01-01 07:51:12	NA	45	Thu	Jan
2	01/01/2015	2015-01-01 08:02:13	NA	60	Thu	Jan
4	01/01/2015	2015-01-01 08:09:12	NA	60	Thu	Jan
5	01/01/2015	2015-01-01 08:16:12	NA	60	Thu	Jan
7	01/01/2015	2015-01-01 08:23:12	NA	60	Thu	Jan
8	01/01/2015	2015-01-01 08:29:12	NA	60	Thu	Jan

A. The increment of time are in multiples of 5's. The types of values that appear, Minimum time is -999.000, 25% of the values are < 45, 50% of the values are < 70, 75% of the values are 95, Maximum time is 300 and the mean is -ve (-2.064) probably due to manual coding or missisng values The different wait time options are seven, Minimum, 1st quater, Median, Mean, 3rd Quater, Maximum and NA'S. B. The NA values in "SPOSTMIN" are, 7635

Question 3

In [9]: `head(myDF)`

	date	datetime	SACTMIN	SPOSTMIN
1	01/01/2015	2015-01-01 07:51:12	NA	45
2	01/01/2015	2015-01-01 08:02:13	NA	60
3	01/01/2015	2015-01-01 08:05:30	54	NA
4	01/01/2015	2015-01-01 08:09:12	NA	60
5	01/01/2015	2015-01-01 08:16:12	NA	60
6	01/01/2015	2015-01-01 08:22:16	55	NA

In [22]: `tail(myDF)`

	date	datetime	SACTMIN	SPOSTMIN
321626	12/28/2021	2021-12-28 22:36:08	NA	45
321627	12/28/2021	2021-12-28 22:42:15	NA	45
321628	12/28/2021	2021-12-28 22:48:12	NA	45
321629	12/28/2021	2021-12-28 22:54:10	NA	45
321630	12/28/2021	2021-12-28 22:54:28	10	NA
321631	12/28/2021	2021-12-28 22:57:34	NA	35

```
In [13]: length(myDF$date[myDF$date == "12/25/2021"])
[1] 156

In [14]: mean(myDF$SACTMIN[myDF$date == "12/25/2021"])
[1] NA

In [15]: mean((myDF$SACTMIN[myDF$date == "12/25/2021"]), na.rm=TRUE)
[1] 48

In [16]: mean((myDF$SACTMIN[myDF$date == "07/26/2021"]), na.rm=TRUE)
[1] 39.6

In [17]: # summer in 2021
mean((myDF$SACTMIN[myDF$date == "08/10/2021"]), na.rm=TRUE)
[1] 43.6

In [18]: #Labour day holiday , 2021
mean((myDF$SACTMIN[myDF$date == "09/06/2021"]), na.rm=TRUE)
[1] 23.25

In [20]: tail(sort(table(myDF$date)), n = 10)
10/31/2021 10/17/2021 10/20/2021 10/05/2021 10/10/2021 10/15/2021 10/29/2021
      323      327      327      329      330      332      333
11/09/2021 10/01/2021 11/08/2021
      344      359      363

In [11]: head(sort(table(myDF$date), decreasing = TRUE), n = 10)
11/08/2021 10/01/2021 11/09/2021 10/29/2021 10/15/2021 10/10/2021 10/05/2021
      363      359      344      333      332      330      329
10/17/2021 10/20/2021 10/08/2021
      327      327      323
```

On Christmas day, the average waiting time is 48. On July 26th, the average waiting time is 39.6. There is a difference between the wait times in the summer and the holidays. The most entries occur on 11/08/2021 on the data set.

Question 4

```
In [46]: dim(myDF)
[1] 321631      4

In [47]: dim(newDF)
[1] 7635      4

In [48]: length(myDF$SPOSTMIN)
[1] 321631

In [49]: length(newDF$SPOSTMIN)
```

```
[1] 7635
```

In [51]: `head(myDF$SPOSTMIN)`

```
[1] 45 60 NA 60 60 NA
```

In [50]: `head(newDF$SPOSTMIN)`

```
[1] NA NA NA NA NA NA
```

In [51]: `head(myDF$SPOSTMIN) + head(newDF$SPOSTMIN)`

```
[1] NA NA NA NA NA NA
```

In [52]: `head(myDF$SPOSTMIN/60) + head(newDF$SPOSTMIN/60)`

```
[1] NA NA NA NA NA NA
```

In [53]: `# divide the whole syntax by 60.
we put the head because we want to see the whole values.
(head(myDF$SPOSTMIN) + head(newDF$SPOSTMIN))/60`

```
[1] NA NA NA NA NA NA
```

In [54]: `myhours <- (myDF$SPOSTMIN + newDF$SPOSTMIN)/60`

Warning message in myDF\$SPOSTMIN + newDF\$SPOSTMIN:
"longer object length is not a multiple of shorter object length"

In [55]: `head(myhours)`

```
[1] NA NA NA NA NA NA
```

In [61]: `length(myhours)`

```
[1] 321631
```

In [40]: `myDF$SPOSTMIN[313997]/60`

```
[1] 0.25
```

In [75]: `newDF$SPOSTMIN[1]/60 # this vector is out of data , so we will
start over again at the begining.
this is called recycling.`

```
[1] 0.75
```

In [56]: `myhours[313997]`

```
[1] NA
```

In [79]: `a <- c(7, 10, 19, 2, 30, 2, 100, 1, 99, 4, 17)`

In [80]: `b <- c(2, 8, 1, 2, 19, 3)`

In [11]: `a + b #warning because not same Length, not multiples`

```
[1] 9 18 20 4 49 5 102 9 100 6 36 13
```

In [1]: `a <- c(7, 10, 19, 2, 30, 2, 100, 1, 99, 4, 17, 10)`

```
In [2]: b <- c(2, 8, 1, 2, 19, 3)
```

```
In [12]: a + b # no warnings because b set is a multiple of a set
```

```
[1] 9 18 20 4 49 5 102 9 100 6 36 13
```

```
In [ ]: # this is called recycling,  
#because (in this case) the values of b keep getting recycled over  
# and over again as long as needed  
# in the examples myDF and newDF, from location 313997 onwards,  
#the values of newDF$SPOSTMIN are just recycled from the start again.
```

The lengths of the column "SPOSTMIN" in the myDF and newDF were 321631 & 313996. The length of the new vector myhours is 321631. At row 313997, newDF\$SPOSTMIN was out of data and shows NA. So we started over again because of recycling.

Question 5

```
In [57]: library(lubridate)
```

```
Warning message in system("timedatectl", intern = TRUE):  
"running command 'timedatectl' had status 1"
```

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

```
The following objects are masked from 'package:base':
```

```
date, intersect, setdiff, union
```

```
In [58]: myDF$weekday <- wday(myDF$datetime, label=TRUE)
```

```
Warning message:  
"tz(): Don't know how to compute timezone for object of class factor; returning "UT  
C". This warning will become an error in the next major version of lubridate."
```

```
In [16]: head(myDF)
```

	date	datetime	SACTMIN	SPOSTMIN	weekday
1	01/01/2015	2015-01-01 07:51:12	NA	45	Thu
2	01/01/2015	2015-01-01 08:02:13	NA	60	Thu
3	01/01/2015	2015-01-01 08:05:30	54	NA	Thu
4	01/01/2015	2015-01-01 08:09:12	NA	60	Thu
5	01/01/2015	2015-01-01 08:16:12	NA	60	Thu
6	01/01/2015	2015-01-01 08:22:16	55	NA	Thu

```
In [17]: tail(myDF)
```

	date	datetime	SACTMIN	SPOSTMIN	weekday
321626	12/28/2021	2021-12-28 22:36:08	NA	45	Tue
321627	12/28/2021	2021-12-28 22:42:15	NA	45	Tue
321628	12/28/2021	2021-12-28 22:48:12	NA	45	Tue
321629	12/28/2021	2021-12-28 22:54:10	NA	45	Tue
321630	12/28/2021	2021-12-28 22:54:28	10	NA	Tue
321631	12/28/2021	2021-12-28 22:57:34	NA	35	Tue

```
In [59]: mean(myDF$SACTMIN[myDF$weekday == "Mon"])
# there are some NA values, we can throw it away.
```

```
[1] NA
```

```
In [22]: mean(myDF$SACTMIN[myDF$weekday == "Mon"], na.rm=TRUE)
[1] -72.86412
```

```
In [23]: head(sort(myDF$SACTMIN))
[1] -92918      0      0      0      0      0
```

```
In [24]: mean(myDF$SACTMIN[myDF$weekday == "Tue"], na.rm=TRUE)
[1] 34.55502
```

```
In [60]: mean(myDF$SACTMIN[myDF$weekday == "Wed"], na.rm=TRUE)
[1] 36.44053
```

```
In [33]: # monday is the day of the week when we are expect to
# have the longest waiting time for the ride.

mean(myDF$SACTMIN[(myDF$weekday == "Mon") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
# this help to eliminate the negation.
```

```
[1] 37.92959
```

```
In [27]: mean(myDF$SACTMIN[(myDF$weekday == "Tue") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
[1] 34.55502
```

```
In [32]: mean(myDF$SACTMIN[(myDF$weekday == "Wed") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
[1] 36.44053
```

```
In [36]: mean(myDF$SACTMIN[(myDF$weekday == "Thu") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
[1] 36.17215
```

```
In [37]: mean(myDF$SACTMIN[(myDF$weekday == "Fri") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
[1] 36.12491
```

```
In [38]: mean(myDF$SACTMIN[(myDF$weekday == "Sat") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
[1] 37.73216
```

```
In [39]: mean(myDF$SACTMIN[(myDF$weekday == "Sun") & (myDF$SACTMIN != -92918)], na.rm=TRUE)
[1] 34.90372
```

```
In [ ]: # tapply function, needs three pieces of data.
# 1. needs the data you want to work on
```

```
# 2. needs to know how you want to break the data into groups
# 3. needs to know what you want to do to each group of the data
# note: any options for that thing you want to do to the data can be given at the end
```

In [40]: `tapply(myDF$SACTMIN, myDF$weekday, mean)`

Sun	Mon	Tue	Wed	Thu	Fri	Sat
NA						

In [41]: `tapply(myDF$SACTMIN, myDF$weekday, mean, na.rm=TRUE)`

Sun	Mon	Tue	Wed	Thu	Fri	Sat
34.90372	-72.86412	34.55502	36.44053	36.17215	36.12491	37.73216

In []: `# we can fix the Monday values this way : by putting this
restriction on all of our data. myDF$SACTMIN !=-92918`

In [61]: `tapply(myDF$SACTMIN, myDF$weekday, mean, na.rm=TRUE)`

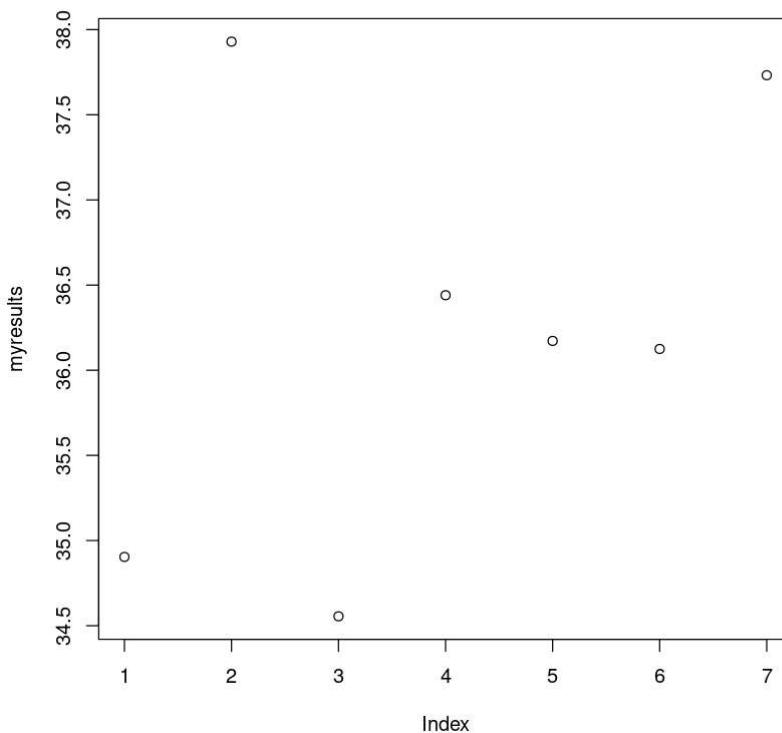
Sun	Mon	Tue	Wed	Thu	Fri	Sat
34.90372	-72.86412	34.55502	36.44053	36.17215	36.12491	37.73216

In [64]: `tapply(myDF$SACTMIN[myDF$SACTMIN !=-92918], myDF$weekday[myDF$SACTMIN !=-92918], mean,`

Sun	Mon	Tue	Wed	Thu	Fri	Sat
34.90372	37.92959	34.55502	36.44053	36.17215	36.12491	37.73216

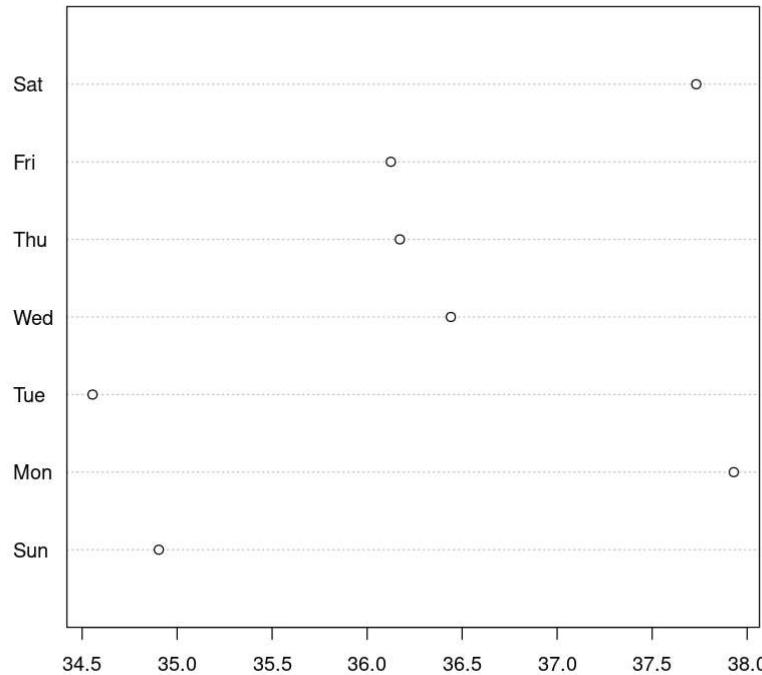
In [65]: `myresults <- tapply(myDF$SACTMIN[myDF$SACTMIN !=-92918],
myDF$weekday[myDF$SACTMIN !=-92918],
mean, na.rm=TRUE)`

In [66]: `plot(myresults)`



In [67]: `dotchart(myresults)`

Warning message in `dotchart(myresults)`:
 "'x' is neither a vector nor a matrix: using `as.numeric(x)`"



In [69]: `myDF$month <- month(myDF$datetime, label=TRUE)`

Warning message:
 "tz(): Don't know how to compute timezone for object of class factor; returning "UTC". This warning will become an error in the next major version of lubridate."

In [70]: `head(myDF)`

	date	datetime	SACTMIN	SPOSTMIN	weekday	month
1	01/01/2015	2015-01-01 07:51:12	NA	45	Thu	Jan
2	01/01/2015	2015-01-01 08:02:13	NA	60	Thu	Jan
3	01/01/2015	2015-01-01 08:05:30	54	NA	Thu	Jan
4	01/01/2015	2015-01-01 08:09:12	NA	60	Thu	Jan
5	01/01/2015	2015-01-01 08:16:12	NA	60	Thu	Jan
6	01/01/2015	2015-01-01 08:22:16	55	NA	Thu	Jan

In [71]: `tail(myDF)`

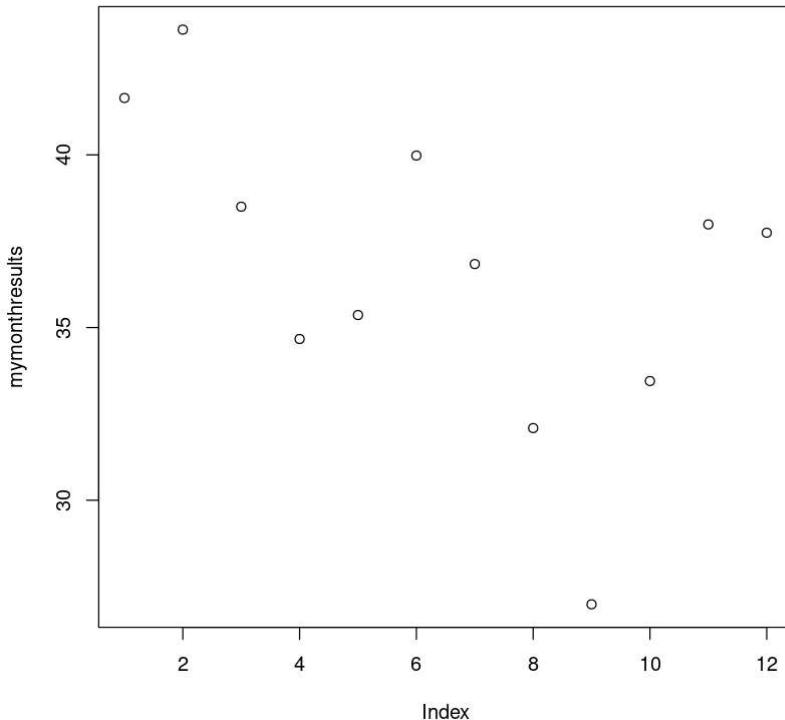
	date	datetime	SACTMIN	SPOSTMIN	weekday	month
321626	12/28/2021	2021-12-28 22:36:08	NA	45	Tue	Dec
321627	12/28/2021	2021-12-28 22:42:15	NA	45	Tue	Dec
321628	12/28/2021	2021-12-28 22:48:12	NA	45	Tue	Dec
321629	12/28/2021	2021-12-28 22:54:10	NA	45	Tue	Dec
321630	12/28/2021	2021-12-28 22:54:28	10	NA	Tue	Dec
321631	12/28/2021	2021-12-28 22:57:34	NA	35	Tue	Dec

In [72]: `tapply(myDF$SACTMIN[myDF$SACTMIN != -92918],
 myDF$month[myDF$SACTMIN != -92918], mean,
 na.rm=TRUE)`

```
Jan      Feb      Mar      Apr      May      Jun      Jul      Aug  
41.64757 43.62733 38.50000 34.66873 35.36293 39.98100 36.83942 32.08961  
Sep      Oct      Nov      Dec  
26.98632 33.45165 37.98672 37.74623
```

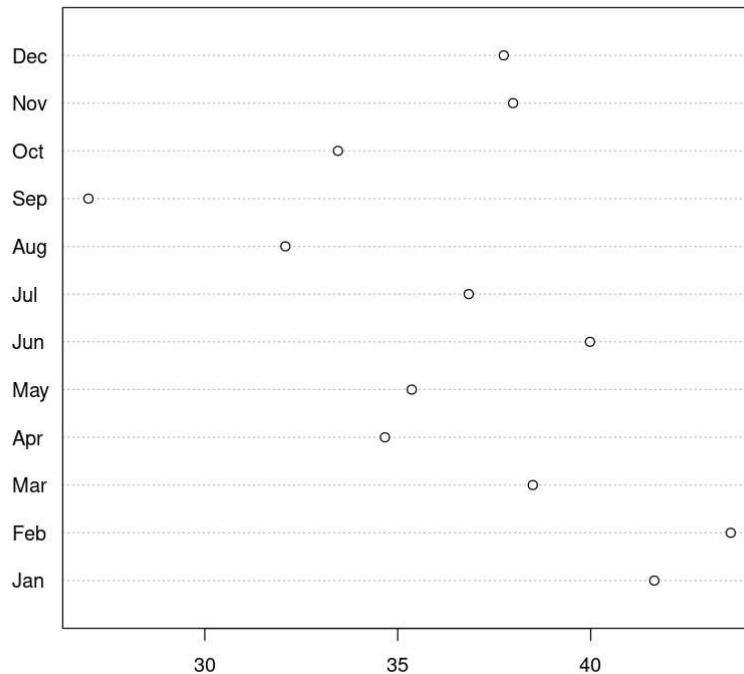
```
In [73]: mymonthresults <- tapply(myDF$SACTMIN[myDF$SACTMIN != -92918],  
                           myDF$month[myDF$SACTMIN != -92918], mean,  
                           na.rm=TRUE)
```

```
In [74]: plot(mymonthresults)
```



```
In [75]: dotchart(mymonthresults)
```

Warning message in dotchart(mymonthresults):
" 'x' is neither a vector nor a matrix: using as.numeric(x)"



Monday is the day of the week in myDF that has the longest average wait times, 37.92959. The dotchart illustrates the data for the average wait times better because it is the days of the week and it is easily traceable. February is the month of the year in myDF that has the longest average wait times, 43.62733.

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.