

# Project 5 Solutions

(Abhimanyu Agarwal)

Collaborators: N/A

TA help:

1) Melissa : Helped me go through Question 3 and 4.

Online resources used: N/A

Question 1

```
#Loads into dataframe called "accidents" using read.csv() and rbind()
accidents <- rbind(read.csv("/class/datamine/data/fars/1975/ACCIDENT.CSV"),
  read.csv("/class/datamine/data/fars/1976/ACCIDENT.CSV"),
  read.csv("/class/datamine/data/fars/1977/ACCIDENT.CSV"),
  read.csv("/class/datamine/data/fars/1978/ACCIDENT.CSV"),
  read.csv("/class/datamine/data/fars/1979/ACCIDENT.CSV"),
  read.csv("/class/datamine/data/fars/1980/ACCIDENT.CSV"),
  read.csv("/class/datamine/data/fars/1981/ACCIDENT.CSV"))

#Reports the dataframes dimensions
dim(accidents)

[1] 300059      45

#Adds a new year column that contains the full year
accidents$YEAR <- factor(paste0("19", accidents$YEAR))

#Prints the updated dataframe
head(accidents)
```

	STATE	COUNTY	MONTH	DAY	YEAR	HOUR	MINUTE	VE_FORMS	PERSONS	VEHICLES
1	1	95	1	4	1975	22	24	3	3	NA
2	1	95	1	2	1975	4	30	1	2	NA
3	1	35	1	1	1975	4	0	1	1	NA
4	1	77	1	2	1975	14	50	1	3	NA
5	1	97	1	5	1975	20	0	1	1	NA
6	1	65	1	4	1975	15	0	2	4	NA

	LAND_USE	CL_TWAY	ROAD_FNC	TA_1_CL	SP_JUR	HARM_EV	MAN_COLL	REL_JUNC
1	1	3	NA	NA	0	12	2	1
2	2	4	NA	NA	0	18	0	1
3	2	6	NA	NA	0	1	0	1
4	2	6	NA	NA	0	27	0	1
5	1	7	NA	NA	0	28	0	1
6	2	4	NA	NA	0	12	4	5

	REL_ROAD	ROAD_FLO	NO_LANES	SP_LIMIT	ALIGNMNT	PROFILE	PAVE_TYP	SUR_COND
1	1	3	4	55	1	1	2	2
2	5	4	2	50	2	1	2	1
3	4	4	2	55	1	2	2	1

```

4      4      4      2      55      2      1      2      1
5      2      4      2      35      1      1      1      1
6      1      4      2      55      1      2      2      1
  TRA_CONT LGT_COND WEATHER HIT_RUN C_M_ZONE NOT_HOUR NOT_MIN ARR_HOUR
1         0         2         1         0        NA        99        99        22
2         0         2         1         0        NA        99        99         4
3         0         2         1         0        NA        99        99         5
4         0         1         1         0        NA        99        99        17
5         0         2         1         0        NA        99        99        20
6         0         1         1         0        NA        99        99        15
  ARR_MIN SCH_BUS CF1 CF2 CF3 FATALS DAY_WEEK DRUNK_DR ST_CASE CITY
1       40      NA  0  0  0         1         7         0  10001  400
2       50      NA  0  0  0         1         5         0  10002  110
3       40      NA  0  0  0         1         4         0  10003   0
4       20      NA  0  0  0         1         5         0  10004   0
5       10      NA  0  0  0         1         1         0  10005 2100
6       15      NA  0  0  0         1         7         0  10006   0
  RAIL
1 *****
2 *****
3 *****
4 *****
5 *****
6 *****

```

```

#Displays the output for the unique()
#Basically displays the unique levels/states
unique(accidents$YEAR)

```

```

[1] 1975 1999 1976 1977 1978 1979 1980 1981
Levels: 1975 1976 1977 1978 1979 1980 1981 1999

```

## Question 2

```
table(accidents$DRUNK_DR)
```

```

      0      1      2      3      4      6
200203 94326 5429  96    4    1

```

```
table(accidents$SCH_BUS, useNA = "always")
```

```

      0      1  <NA>
220306  845 78908

```

```

#Gives you a table to see when is that true, false or N/A
table((accidents$DRUNK_DR != 0)&(accidents$SCH_BUS == 1), useNA = "always")

```

```

  FALSE  TRUE  <NA>
284228   101 15730

```

```

#Number of accidents in totality with atleast 1 drunk driver and 1 bus.
length(which((accidents$DRUNK_DR != 0)&(accidents$SCH_BUS == 1)))

```

```
[1] 101
```

```
#There are 101 accidents with at least 1 drunk driver with a school bus.
```

### Question 3

```
#Tells how many accidents took place in the given span of time (for each year).
```

```
table(accidents$YEAR[(accidents$DRUNK_DR != 0)&(accidents$SCH_BUS == 1)])
```

```
1975 1976 1977 1978 1979 1980 1981 1999
    0    0   12   37   17   12   23    0
```

```
#Tells what is maximum amount of accidents.
```

```
max(table(accidents$YEAR[(accidents$DRUNK_DR != 0)&(accidents$SCH_BUS == 1)]))
```

```
[1] 37
```

```
#Tells which year witnessed maximum amount of accidents.
```

```
which.max(table(accidents$YEAR[(accidents$DRUNK_DR != 0)&(accidents$SCH_BUS == 1)]))
```

```
1978
```

```
4
```

```
#1978 has the maximum number of accidents
```

### Question 4

```
#Groups the variable/ first argument into categories based on the second argument
```

```
#Applies specific mathematical function like variance, sum etc to the grouped data.
```

```
tapply(accidents$PERSONS, accidents$DRUNK_DR, mean)
```

```
      0      1      2      3      4      6
2.615540 2.474079 3.660711 5.197917 5.250000 6.000000
```

```
###Question 5
```

```
#Captures the relevant states data into the dataframe "state"
```

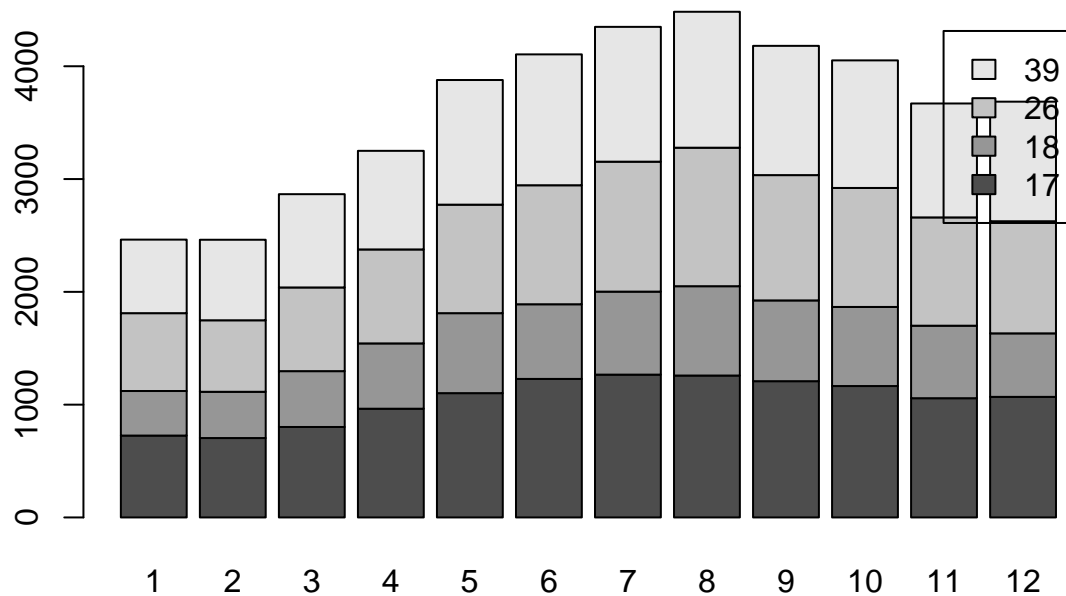
```
state <- accidents[(accidents$STATE %in% c(17,18,26,39)), ]
```

```
#Table t is formulated
```

```
t <- table(state$STATE, state$MONTH)
```

```
#Barplot with legend is generated
```

```
barplot(t, legend = rownames(t))
```



*#From the bar graph, it can be concluded that the summer months (6 to 9) witness  
 #the greatest rise in the number of accidents. As the summer passes a decrease  
 #in the incidents can be observed. The increment in the accident numbers can be  
 #because of people could be on vacations during the summer or partying with friends.  
 #With all the hustle, maybe people just aren't as attentive as they should be.*

Submitting deliverables: project05.RMD, project05.R and project05.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.