

# Project 5 --Seyi Ogunmodede

**Instructor:** Dr. Mark Daniel John Smith, Alice Jones

- Help with videos on how to write a function.

**Collaboration:**

## Question 1

```
In [2]: options(jupyter.rich_display = F)
```

```
In [3]: options(repr.matrix.max.cols=25, repr.matrix.max.rows=200)
```

```
In [4]: myDF <- read.csv("/anvil/projects/tdm/data/election/escaped/itcont2020sample.txt",
                     sep="|")
```

```
In [5]: head(myDF)
```

CMTE_ID	AMNDT_IND	RPT_TP	TRANSACTION_PGI	IMAGE_NUM	TRANSACTION_TP
1 C00401224	N	YE		2.020013e+17	24T
2 C00401224	N	M4	P	2.020042e+17	24T
3 C00193433	N	YE	P	2.020012e+17	15
4 C00401224	N	YE		2.020013e+17	24T
5 C00401224	N	MY		2.019073e+17	24T
6 C00618371	N	Q3	P	2.019102e+17	15
	ENTITY_TP	NAME	CITY	STATE	ZIP_CODE EMPLOYER
1 IND	SOLOMON, VICTOR	CAPE CORAL	FL	33904	RETIRED
2 IND	MYERS, ELIZABETH	SCOTTSVILLE	NY	14546	NOT EMPLOYED
3 IND	HASKINS, GEORGE	HARTSDALE	NY	10530	NOT EMPLOYED
4 IND	KOLBE, NINA	WASHINGTON	DC	20003	SELF
5 IND	ESSENBERG, PENNY	GRAND RAPIDS	MI	49534	
6 IND	GIEBEL, BETTY	KISSIMMEE	FL	34744	HOMEMAKER
	OCCUPATION		TRANSACTION_DT	TRANSACTION_AMT	OTHER_ID TRAN_ID
1 NOT EMPLOYED		7112019	15		C00042366 SA11AI_165851302
2 NOT EMPLOYED		3042020	39		C00696948 SA11AI_216676150
3 NOT EMPLOYED		12182019	15		6260719
4 HEALTH PROFESSIONAL		8302019	10		C00637074 SA11AI_174544101
5		3212019	16		C00580068 SB28A_145393165
6 HOMEMAKER		7112019	42		SA11AI.178611
	FILE_NUM	MEMO_CD			
1 1378435					
2 1402724					
3 1371576					
4 1378435					
5 1344765					
6 1358506					
	MEMO_TEXT				
1 EARMARKED FOR DSCC (C00042366)					
2 EARMARKED FOR BERNIE 2020 (C00696948)					
3					
4 EARMARKED FOR FINKENAUER FOR CONGRESS (C00637074)					
5 REFUND OF CONTRIBUTION, INITIALLY EARMARKED FOR PROGRESSIVE TURNOUT PROJECT (C00580068)					
6					
	SUB_ID				
1 4.03022e+18					
2 4.05142e+18					
3 4.01222e+18					
4 4.03022e+18					
5 4.08282e+18					
6 4.10182e+18					

```
In [6]: class(myDF$ TRANSACTION_DT)
[1] "integer"
```

```
In [7]: 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 [8]: library(lubridate, warn.conflicts = FALSE)
```

```
In [9]: library(lubridate)
```

```
In [10]: dim(myDF)
```

```
[1] 4678762      21
```

```
In [11]: class(head(myDF$TRANSACTION_DT))
```

```
[1] "integer"
```

```
In [12]: head(myDF$TRANSACTION_DT)
```

```
[1] 7112019 3042020 12182019 8302019 3212019 7112019
```

```
In [13]: mdy(head(myDF$TRANSACTION_DT))
```

```
[1] "2019-07-11" "2020-03-04" "2019-12-18" "2019-08-30" "2019-03-21"  
[6] "2019-07-11"
```

```
In [14]: class(mdy(head(myDF$TRANSACTION_DT)))  
# with mdy the class is date, while before it was integer.
```

```
[1] "Date"
```

```
In [15]: myDF$newdates <- mdy(myDF$TRANSACTION_DT)
```

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

CMTE_ID	AMNDT_IND	RPT_TP	TRANSACTION_PGI	IMAGE_NUM	TRANSACTION_TP
1 C00401224	N	YE		2.020013e+17	24T
2 C00401224	N	M4	P	2.020042e+17	24T
3 C00193433	N	YE	P	2.020012e+17	15
4 C00401224	N	YE		2.020013e+17	24T
5 C00401224	N	MY		2.019073e+17	24T
6 C00618371	N	Q3	P	2.019102e+17	15
	ENTITY_TP	NAME	CITY	STATE	ZIP_CODE EMPLOYER
1 IND	SOLOMON, VICTOR	CAPE CORAL	FL	33904	RETIRED
2 IND	MYERS, ELIZABETH	SCOTTSVILLE	NY	14546	NOT EMPLOYED
3 IND	HASKINS, GEORGE	HARTSDALE	NY	10530	NOT EMPLOYED
4 IND	KOLBE, NINA	WASHINGTON	DC	20003	SELF
5 IND	ESSENBERG, PENNY	GRAND RAPIDS	MI	49534	
6 IND	GIEBEL, BETTY	KISSIMMEE	FL	34744	HOMEMAKER
	OCCUPATION		TRANSACTION_DT	TRANSACTION_AMT	OTHER_ID TRAN_ID
1 NOT EMPLOYED			7112019	15	C00042366 SA11AI_165851302
2 NOT EMPLOYED			3042020	39	C00696948 SA11AI_216676150
3 NOT EMPLOYED			12182019	15	6260719
4 HEALTH PROFESSIONAL			8302019	10	C00637074 SA11AI_174544101
5			3212019	16	C00580068 SB28A_145393165
6 HOMEMAKER			7112019	42	SA11AI.178611
	FILE_NUM	MEMO_CD			
1 1378435					
2 1402724					
3 1371576					
4 1378435					
5 1344765					
6 1358506		MEMO_TEXT			
1 EARMARKED FOR DSCC (C00042366)					
2 EARMARKED FOR BERNIE 2020 (C00696948)					
3					
4 EARMARKED FOR FINKENAUER FOR CONGRESS (C00637074)					
5 REFUND OF CONTRIBUTION, INITIALLY EARMARKED FOR PROGRESSIVE TURNOUT PROJECT (C00580068)					
6					
	SUB_ID	newdates			
1 4.03022e+18		2019-07-11			
2 4.05142e+18		2020-03-04			
3 4.01222e+18		2019-12-18			
4 4.03022e+18		2019-08-30			
5 4.08282e+18		2019-03-21			
6 4.10182e+18		2019-07-11			

```
In [ ]: # Using tapply, add the values in the TRANSACTION_AMT column,
# according to the values in the newdate column.
```

```
In [ ]: # tapply takes 3 inputs:
# the data to work on
# the way to break the date into groups
# the function to run on each piece of data
```

```
In [20]: sum(myDF$TRANSACTION_AMT)
```

```
[1] 1067589940
```

```
In [19]: tapply(myDF$TRANSACTION_AMT, myDF$newdates, sum)
# the way Tapply work, you put the function at the end,
```

*# i want to sum up the variables at the beginning, the TRANSACTION\_AMT  
# not all but broken into groups according to the newdates*

2002-02-13	2010-08-08	2015-06-28	2015-06-29	2016-06-03	2016-06-04	2017-11-20
500	500	250	275	1000	250	850
2017-12-08	2018-01-06	2018-01-20	2018-01-24	2018-02-03	2018-02-11	2018-02-16
1700	850	850	850	850	850	850
2018-02-19	2018-03-04	2018-03-22	2018-03-31	2018-04-07	2018-04-13	2018-04-27
850	300	850	3090	850	850	850
2018-04-30	2018-05-01	2018-05-03	2018-05-04	2018-05-10	2018-05-18	2018-05-21
850	150	850	850	850	850	1150
2018-05-24	2018-05-25	2018-05-31	2018-06-14	2018-06-26	2018-07-22	2018-07-31
850	150	250	150	2800	850	850
2018-08-06	2018-08-12	2018-08-24	2018-09-01	2018-09-10	2018-09-15	2018-09-26
1960	850	2550	850	325	850	850
2018-10-02	2018-10-05	2018-10-07	2018-10-10	2018-10-23	2018-10-25	2018-11-03
345	850	142	100	250	850	850
2018-11-05	2018-11-06	2018-11-07	2018-11-09	2018-11-11	2018-11-14	2018-11-15
850	2	850	56800	15	1150	9469
2018-11-16	2018-11-17	2018-11-18	2018-11-19	2018-11-20	2018-11-21	2018-11-22
10726	3300	500	25400	200	2200	7300
2018-11-23	2018-11-25	2018-11-27	2018-11-28	2018-11-29	2018-12-01	2018-12-02
175	25	124	376	750	500	250
2018-12-03	2018-12-05	2018-12-07	2018-12-09	2018-12-10	2018-12-11	2018-12-12
5750	250	9	-497	1025	10900	312
2018-12-13	2018-12-14	2018-12-16	2018-12-17	2018-12-18	2018-12-19	2018-12-20
3100	630	170	2384	3242	476	1606
2018-12-21	2018-12-23	2018-12-24	2018-12-25	2018-12-26	2018-12-27	2018-12-28
725	25	187	122	553	1222	4065
2018-12-29	2018-12-30	2018-12-31	2019-01-01	2019-01-02	2019-01-03	2019-01-04
325	6925	19217	286355	224173	343765	285968
2019-01-05	2019-01-06	2019-01-07	2019-01-08	2019-01-09	2019-01-10	2019-01-11
130386	142742	283726	345262	388761	528628	489168
2019-01-12	2019-01-13	2019-01-14	2019-01-15	2019-01-16	2019-01-17	2019-01-18
165037	143196	413388	514494	453520	568698	482521
2019-01-19	2019-01-20	2019-01-21	2019-01-22	2019-01-23	2019-01-24	2019-01-25
99767	165167	365924	531936	622508	467723	734624
2019-01-26	2019-01-27	2019-01-28	2019-01-29	2019-01-30	2019-01-31	2019-02-01
151261	445038	806948	735979	848832	1831005	1719995
2019-02-02	2019-02-03	2019-02-04	2019-02-05	2019-02-06	2019-02-07	2019-02-08
221447	268864	709790	1713867	716363	947635	789129
2019-02-09	2019-02-10	2019-02-11	2019-02-12	2019-02-13	2019-02-14	2019-02-15
269338	440437	820580	1092709	901121	827222	987777
2019-02-16	2019-02-17	2019-02-18	2019-02-19	2019-02-20	2019-02-21	2019-02-22
232636	311150	387157	1825044	1017885	1063580	1088359
2019-02-23	2019-02-24	2019-02-25	2019-02-26	2019-02-27	2019-02-28	2019-03-01
213831	300114	1039855	877042	993708	2542423	2102897
2019-03-02	2019-03-03	2019-03-04	2019-03-05	2019-03-06	2019-03-07	2019-03-08
304253	399438	1249647	1013727	1214972	877371	1365114
2019-03-09	2019-03-10	2019-03-11	2019-03-12	2019-03-13	2019-03-14	2019-03-15
349163	390423	964548	1491613	1058149	2292228	2006153
2019-03-16	2019-03-17	2019-03-18	2019-03-19	2019-03-20	2019-03-21	2019-03-22
341329	473592	1607241	1691067	1679980	1202932	1771741
2019-03-23	2019-03-24	2019-03-25	2019-03-26	2019-03-27	2019-03-28	2019-03-29
410224	526993	2027369	2039442	2180524	1941410	3811357
2019-03-30	2019-03-31	2019-04-01	2019-04-02	2019-04-03	2019-04-04	2019-04-05
1526737	4369255	2408583	985203	829230	977438	1072370
2019-04-06	2019-04-07	2019-04-08	2019-04-09	2019-04-10	2019-04-11	2019-04-12
276650	315657	1585431	934592	1063696	1236980	1044288
2019-04-13	2019-04-14	2019-04-15	2019-04-16	2019-04-17	2019-04-18	2019-04-19
264899	586001	1328129	1013334	1477031	1183482	1423080
2019-04-20	2019-04-21	2019-04-22	2019-04-23	2019-04-24	2019-04-25	2019-04-26
330465	284249	1020372	957776	2626638	2520110	1638413

2019-04-27	2019-04-28	2019-04-29	2019-04-30	2019-05-01	2019-05-02	2019-05-03
423090	558746	1551699	2663527	1327223	1611982	2750129
2019-05-04	2019-05-05	2019-05-06	2019-05-07	2019-05-08	2019-05-09	2019-05-10
357642	424795	1261678	1131866	1006466	1425091	1168083
2019-05-11	2019-05-12	2019-05-13	2019-05-14	2019-05-15	2019-05-16	2019-05-17
282666	364190	1322745	1564637	1262397	1207509	1141314
2019-05-18	2019-05-19	2019-05-20	2019-05-21	2019-05-22	2019-05-23	2019-05-24
359465	501392	1400443	2220153	894373	1562807	1387274
2019-05-25	2019-05-26	2019-05-27	2019-05-28	2019-05-29	2019-05-30	2019-05-31
337735	280991	262263	2162133	2519604	1367882	2730410
2019-06-01	2019-06-02	2019-06-03	2019-06-04	2019-06-05	2019-06-06	2019-06-07
420111	358819	1944736	1225206	1114106	1283286	1121565
2019-06-08	2019-06-09	2019-06-10	2019-06-11	2019-06-12	2019-06-13	2019-06-14
329499	373699	1369204	1165822	1050186	2065717	1522541
2019-06-15	2019-06-16	2019-06-17	2019-06-18	2019-06-19	2019-06-20	2019-06-21
471873	341919	1694723	3056274	1421896	2194020	1614011
2019-06-22	2019-06-23	2019-06-24	2019-06-25	2019-06-26	2019-06-27	2019-06-28
469556	468593	1732344	2987319	2016351	2784271	4039173
2019-06-29	2019-06-30	2019-07-01	2019-07-02	2019-07-03	2019-07-04	2019-07-05
1723680	6089037	1670385	1002170	640287	220606	648757
2019-07-06	2019-07-07	2019-07-08	2019-07-09	2019-07-10	2019-07-11	2019-07-12
271992	276550	1723534	2005316	1220522	1248242	973333
2019-07-13	2019-07-14	2019-07-15	2019-07-16	2019-07-17	2019-07-18	2019-07-19
299909	454620	1422993	1054635	902927	1075053	1332860
2019-07-20	2019-07-21	2019-07-22	2019-07-23	2019-07-24	2019-07-25	2019-07-26
436728	449409	1218090	2075894	1153613	1026204	1843576
2019-07-27	2019-07-28	2019-07-29	2019-07-30	2019-07-31	2019-08-01	2019-08-02
422850	596390	1309177	1616703	3276869	1558773	1271089
2019-08-03	2019-08-04	2019-08-05	2019-08-06	2019-08-07	2019-08-08	2019-08-09
384953	408195	1088546	1456662	1260074	1173654	3487541
2019-08-10	2019-08-11	2019-08-12	2019-08-13	2019-08-14	2019-08-15	2019-08-16
451449	488384	1145421	1429993	1277055	1592386	1334483
2019-08-17	2019-08-18	2019-08-19	2019-08-20	2019-08-21	2019-08-22	2019-08-23
431108	519520	1305842	1359566	1054856	1387714	1460490
2019-08-24	2019-08-25	2019-08-26	2019-08-27	2019-08-28	2019-08-29	2019-08-30
433936	496592	1305663	1275489	1653861	1929083	2263077
2019-08-31	2019-09-01	2019-09-02	2019-09-03	2019-09-04	2019-09-05	2019-09-06
1920122	596724	376367	1423570	2343082	1261279	1466176
2019-09-07	2019-09-08	2019-09-09	2019-09-10	2019-09-11	2019-09-12	2019-09-13
625035	561429	2955745	1553557	2576205	1657594	1934019
2019-09-14	2019-09-15	2019-09-16	2019-09-17	2019-09-18	2019-09-19	2019-09-20
644332	778991	1889728	1445201	1879394	3255324	1923679
2019-09-21	2019-09-22	2019-09-23	2019-09-24	2019-09-25	2019-09-26	2019-09-27
732994	911855	2481666	2324698	2372121	3451163	3349036
2019-09-28	2019-09-29	2019-09-30	2019-10-01	2019-10-02	2019-10-03	2019-10-04
1624274	2043461	11248433	1906926	1374001	1577362	1286900
2019-10-05	2019-10-06	2019-10-07	2019-10-08	2019-10-09	2019-10-10	2019-10-11
503794	730158	2103705	2349552	1324080	1444266	1251892
2019-10-12	2019-10-13	2019-10-14	2019-10-15	2019-10-16	2019-10-17	2019-10-18
456656	499498	838354	3302133	1672420	2140214	2233004
2019-10-19	2019-10-20	2019-10-21	2019-10-22	2019-10-23	2019-10-24	2019-10-25
673898	708166	1828871	1405484	1567518	1877222	1963704
2019-10-26	2019-10-27	2019-10-28	2019-10-29	2019-10-30	2019-10-31	2019-11-01
577457	655347	1854755	1485982	2015625	5801325	1482658
2019-11-02	2019-11-03	2019-11-04	2019-11-05	2019-11-06	2019-11-07	2019-11-08
489808	630506	2311178	1616197	1421163	1611042	2707289
2019-11-09	2019-11-10	2019-11-11	2019-11-12	2019-11-13	2019-11-14	2019-11-15
556079	639339	849912	1797754	1510422	1989788	2485904
2019-11-16	2019-11-17	2019-11-18	2019-11-19	2019-11-20	2019-11-21	2019-11-22
699888	762706	1866925	2686134	61029208	2347140	2861556

	2019-11-23	2019-11-24	2019-11-25	2019-11-26	2019-11-27	2019-11-28	2019-11-29
	618050	765050	1872741	1712282	1293407	445850	1328499
	2019-11-30	2019-12-01	2019-12-02	2019-12-03	2019-12-04	2019-12-05	2019-12-06
	2493312	656788	1937288	1615756	2053434	1787557	3620881
	2019-12-07	2019-12-08	2019-12-09	2019-12-10	2019-12-11	2019-12-12	2019-12-13
	609405	740880	1906457	2482822	1682342	1857648	2579446
	2019-12-14	2019-12-15	2019-12-16	2019-12-17	2019-12-18	2019-12-19	2019-12-20
	765171	1068117	2267592	3304527	2763266	4834897	13720524
	2019-12-21	2019-12-22	2019-12-23	2019-12-24	2019-12-25	2019-12-26	2019-12-27
	978626	868032	3997170	2279853	757424	1739673	5522488
	2019-12-28	2019-12-29	2019-12-30	2019-12-31	2020-01-01	2020-01-02	2020-01-03
	1555677	1426405	5640891	12463773	512912	676797	817690
	2020-01-04	2020-01-05	2020-01-06	2020-01-07	2020-01-08	2020-01-09	2020-01-10
	448177	560156	1074375	1482340	2773368	1176143	2302235
	2020-01-11	2020-01-12	2020-01-13	2020-01-14	2020-01-15	2020-01-16	2020-01-17
	691825	578961	1834570	1966775	4378806	1626644	1901076
	2020-01-18	2020-01-19	2020-01-20	2020-01-21	2020-01-22	2020-01-23	2020-01-24
	670463	754740	990074	1796661	2612566	19674963	1691427
	2020-01-25	2020-01-26	2020-01-27	2020-01-28	2020-01-29	2020-01-30	2020-01-31
	865238	842613	2254672	2263028	2302710	2796299	4918652
	2020-02-01	2020-02-02	2020-02-03	2020-02-04	2020-02-05	2020-02-06	2020-02-07
	1165294	914210	3094231	2757217	4069356	4256715	9295265
	2020-02-08	2020-02-09	2020-02-10	2020-02-11	2020-02-12	2020-02-13	2020-02-14
	1535016	1403996	3301579	3204444	4677244	128122042	3880426
	2020-02-15	2020-02-16	2020-02-17	2020-02-18	2020-02-19	2020-02-20	2020-02-21
	1697205	1253001	1463749	5645990	3806077	4620127	3147509
	2020-02-22	2020-02-23	2020-02-24	2020-02-25	2020-02-26	2020-02-27	2020-02-28
	1531281	1654197	4380206	89123625	3583972	3979764	5897681
	2020-02-29	2020-03-01	2020-03-02	2020-03-03	2020-03-04	2020-03-05	2020-03-06
	5012038	2900184	3966423	4854497	5814943	4637253	3876184
	2020-03-07	2020-03-08	2020-03-09	2020-03-10	2020-03-11	2020-03-12	2020-03-13
	1333327	1431096	8149441	3095594	2602841	3703124	2911791
	2020-03-14	2020-03-15	2020-03-16	2020-03-17	2020-03-18	2020-03-19	2020-03-20
	804104	2455542	2234646	1968772	2775655	1952745	2394683
	2020-03-21	2020-03-22	2020-03-23	2020-03-24	2020-03-25	2020-03-26	2020-03-27
	646034	702572	1393339	2311539	1944412	2909706	2152029
	2020-03-28	2020-03-29	2020-03-30	2020-03-31	2020-04-01	2020-04-02	2020-04-03
	1069381	1249111	4271112	10400920	257300	251187	334071
	2020-04-04	2020-04-05	2020-04-06	2020-04-07	2020-04-08	2020-04-09	2020-04-10
	55284	78473	194809	205210	1256284	209528	585818
	2020-04-11	2020-04-12	2020-04-13	2020-04-14	2020-04-15	2020-04-16	2020-04-17
	37948	51158	253328	160274	1549357	255613	1179625
	2020-04-18	2020-04-19	2020-04-20	2020-04-21	2020-04-22	2020-04-23	2020-04-24
	85223	105892	1284372	216202	1640185	187523	794947
	2020-04-25	2020-04-26	2020-04-27	2020-04-28	2020-04-29	2020-04-30	2020-05-01
	46614	85551	492371	414259	219346	2203021	20000
	2020-05-02	2020-05-03	2020-05-04	2020-05-05	2020-05-06	2020-05-07	2020-05-08
	6465	4640	157196	28180	26904	32189	28996
	2020-05-09	2020-05-10	2020-05-11	2020-05-12	2020-05-13	2020-05-14	2020-12-16
	10359	18624	27734	48912	72568	250	500
	2020-12-18						
		2800					

In [24]: `myresults <- tapply(myDF$TRANSACTION_AMT, myDF$newdates, sum)`

In [23]: `head(myresults)`

	2002-02-13	2010-08-08	2015-06-28	2015-06-29	2016-06-03	2016-06-04
	500	500	250	275	1000	250

In [ ]:

```
In [25]: length(sort(unique(myDF$newdates)))
# i needed to sort this to put the dates in order as they would appear on the plot.
# on the x-axis, it has the unique dates listed in sorted order.
# on the y-axis , it has the corresponding values
```

```
[1] 596
```

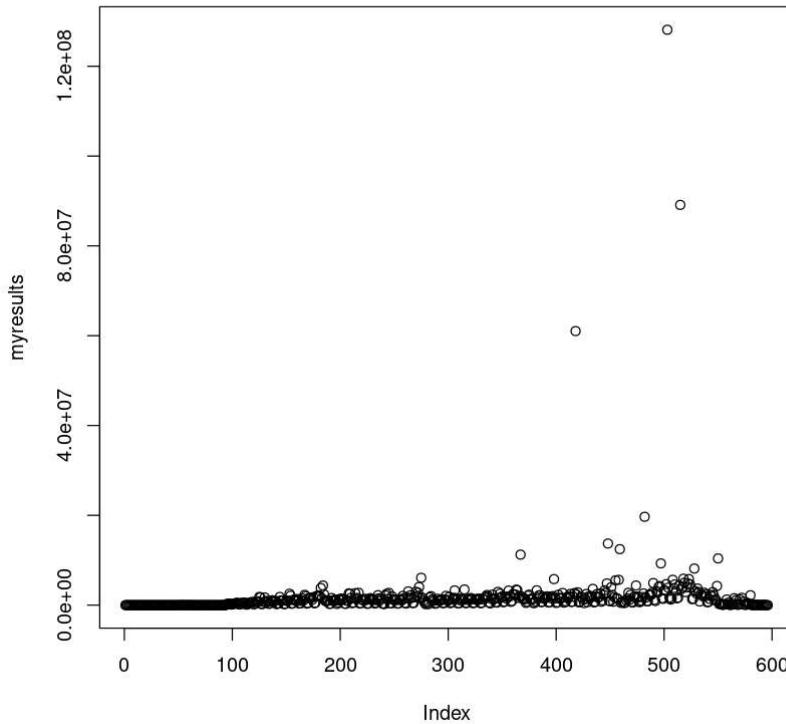
```
In [26]: length(unique(myDF$newdates))
# there is discrepancy, meaning there is NA in there
```

```
[1] 597
```

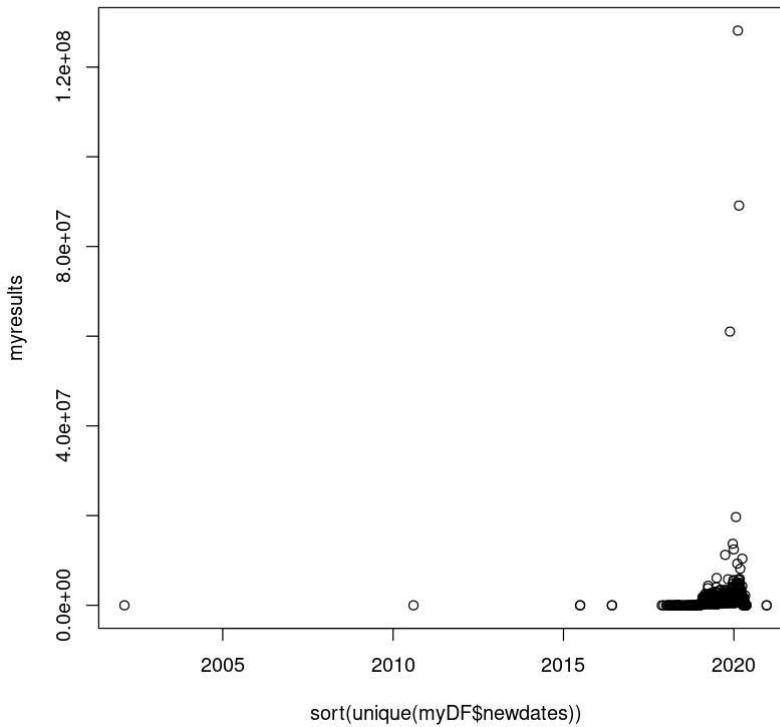
```
In [27]: length(myresults)
```

```
[1] 596
```

```
In [28]: plot(myresults)
```



```
In [29]: plot(sort(unique(myDF$newdates)),myresults)
# This will align the right dollars with the right dates, on the right axes.
```



In [ ]:

Use the mdy function (from the lubridate library) on the column TRANSACTION\_DT, to create a new column named newdates. newdates

2019-07-11 2020-03-04 2019-12-18 2019-08-30 2019-03-21 2019-07-11

Using tapply, add the values in the TRANSACTION\_AMT column, according to the values in the newdate column.

2002-02-13 2010-08-08 2015-06-28 2015-06-29 2016-06-03 2016-06-04 2017-11-20 500 500  
250 275 1000 250 850.....

Plot the dates on the x-axis and the information we found in part b on the y-axis.

## Question 2

```
In [30]: class("01/01/2019")
# i need to convert this to date to be able to
work with the term date not string
```

[1] "character"

```
In [31]: class(mdy("01/01/2019"))
# now converted to date
```

[1] "Date"

```
In [32]: newDF <- subset(myDF, (newdates >= mdy("01/01/2019"))
& (newdates <= mdy("05/15/2019")))
```

```
In [33]: dim(myDF)
```

```
[1] 4678762      22
```

```
In [34]: dim(newDF)
```

```
[1] 495435      22
```

```
In [35]: mynewresults <- tapply(newDF$TRANSACTION_AMT,  
                           newDF$newdates, sum)
```

```
In [36]: head(mynewresults)
```

```
2019-01-01 2019-01-02 2019-01-03 2019-01-04 2019-01-05 2019-01-06  
286355      224173      343765      285968      130386      142742
```

```
In [37]: tail(mynewresults)
```

```
2019-05-10 2019-05-11 2019-05-12 2019-05-13 2019-05-14 2019-05-15  
1168083     282666     364190     1322745    1564637     1262397
```

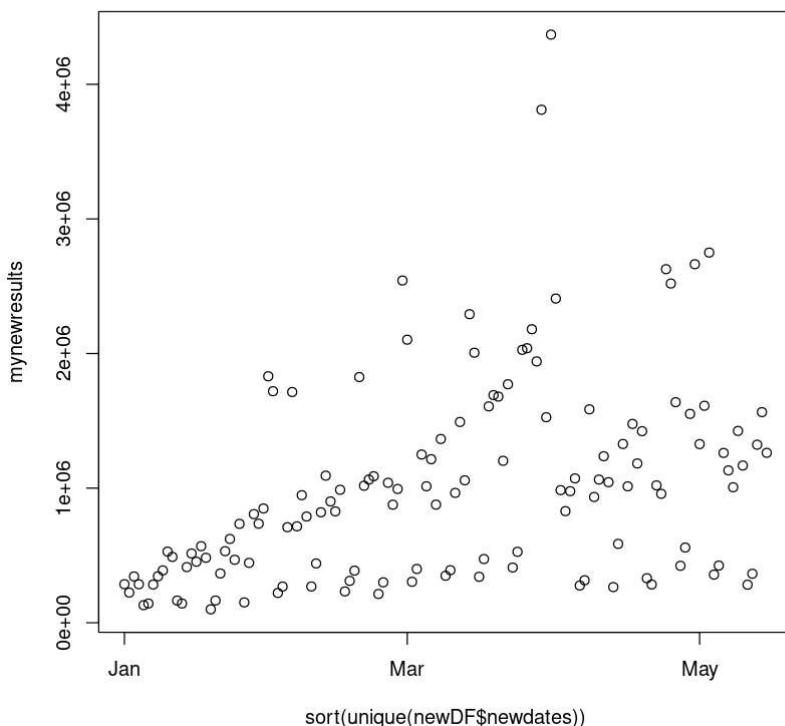
```
In [38]: length(sort(unique(newDF$newdates)))  
# to know how many days are in it.
```

```
[1] 135
```

```
In [39]: length(mynewresults)
```

```
[1] 135
```

```
In [40]: plot(sort(unique(newDF$newdates)),mynewresults)
```



In [ ]:

In [ ]:

Create a new dataframe that only contains data for the dates in the range 01/01/2019-05/15/2019

Plot the new dataframe

What do you notice about the data? From the plot there was steady increase from January, it also shows that more giving occurs on different days of the week, certain days has more donations than the other.

## Question 3

```
In [ ]: # tapply takes 3 inputs:  
# the data to work on  
# the way to break the date into groups  
# the function to run on each piece of data
```

```
In [41]: v <- tapply(myDF$TRANSACTION_AMT, myDF$STATE, sum)
```

```
In [42]: head(v)
```

	AA	AB	AE	AK	AL
431548	32243	1474	87137	1627993	7195054

```
In [43]: tail(v)
```

	WA	WI	WV	WY	YT	ZZ
19736146	5911445	1181879	3554048		162	823468

```
In [44]: length(v)
```

```
[1] 78
```

```
In [45]: v
```

		AA	AB	AE	AK	AL	AP	AR
431548		32243	1474	87137	1627993	7195054	49240	5690116
AS	AZ		BC	CA	CO	CT	DC	DE
2908	11750510		12083	151842431	18025653	12065158	37280905	2120061
FL	FM		GA	GU	HI	IA	ID	IL
44118779		2250	17026567		78087	1730014	4291247	1626126
IN	KS		KY	LA	MA	MB	MD	ME
8906170	4311170		4440149	6202190	32308431		213	14940848
MH	MI		MN	MO	MP	MS	MT	NB
5000	15933204		9990430	7123990	10098	1844410	2389091	417
NC	ND		NE	NH	NJ	NL	NM	NS
13368140	1665152		2418525	2970383	16327791		909	4137547
NT	NV		NY	OH	OK	ON	OR	PA
9	9346089	364883860		15076354	5402240		17743	7088869
PE	PR		PW	QC	RI	SC	SD	SK
50	457693		5000	4347	1636246	5586743	1212866	66
SP	TN		TX	UK	UT	VA	VI	VT
2	9043359	63719103		25	3517485	24995172		86549
WA	WI		WV	WY	YT	ZZ		
19736146	5911445		1181879	3554048		162	823468	

In [46]: `sort(v)`

SP	NT	UK	PE	SK	YT	MB	NB
2	9	25	50	66	162	213	417
NS	NL	AB	FM	AS	QC	MH	PW
678	909	1474	2250	2908	4347	5000	5000
MP	BC	ON	AA	AP	GU	VI	AE
10098	12083	17743	32243	49240	78087	86549	87137
	PR	ZZ	WV	SD	VT	ID	AK
431548	457693	823468	1181879	1212866	1492795	1626126	1627993
RI	ND	HI	MS	DE	MT	NE	NH
1636246	1665152	1730014	1844410	2120061	2389091	2418525	2970383
ME	UT	WY	NM	IA	KS	KY	OK
3393310	3517485	3554048	4137547	4291247	4311170	4440149	5402240
SC	AR	WI	LA	OR	MO	AL	IN
5586743	5690116	5911445	6202190	7088869	7123990	7195054	8906170
TN	NV	MN	AZ	CT	NC	MD	OH
9043359	9346089	9990430	11750510	12065158	13368140	14940848	15076354
MI	NJ	GA	CO	WA	VA	PA	IL
15933204	16327791	17026567	18025653	19736146	24995172	25093386	27940911
MA	DC	FL	TX	CA	NY		
32308431	37280905	44118779	63719103	151842431	364883860		

In [47]: `head(myDF$CITY)`

```
[1] "CAPE CORAL"    "SCOTTSVILLE"   "HARTSDALE"      "WASHINGTON"     "GRAND RAPIDS"
[6] "KISSIMMEE"
```

In [48]: `head(myDF$STATE)`

```
[1] "FL"  "NY"  "NY"  "DC"  "MI"  "FL"
```

In [49]: `paste(head(myDF$CITY), head(myDF$STATE), sep= ", ")`

```
[1] "CAPE CORAL, FL"  "SCOTTSVILLE, NY" "HARTSDALE, NY"   "WASHINGTON, DC"
[5] "GRAND RAPIDS, MI" "KISSIMMEE, FL"
```

In [48]: `myDF$citystatepair <- paste(myDF$CITY, myDF$STATE, sep= ", ")`

In [50]: `head(myDF)`

CMTE_ID	AMNDT_IND	RPT_TP	TRANSACTION_PGI	IMAGE_NUM	TRANSACTION_TP
1 C00401224	N	YE		2.020013e+17	24T
2 C00401224	N	M4	P	2.020042e+17	24T
3 C00193433	N	YE	P	2.020012e+17	15
4 C00401224	N	YE		2.020013e+17	24T
5 C00401224	N	MY		2.019073e+17	24T
6 C00618371	N	Q3	P	2.019102e+17	15
ENTITY_TP	NAME	CITY	STATE	ZIP_CODE	EMPLOYER
1 IND	SOLOMON, VICTOR	CAPE CORAL	FL	33904	RETIRED
2 IND	MYERS, ELIZABETH	SCOTTSVILLE	NY	14546	NOT EMPLOYED
3 IND	HASKINS, GEORGE	HARTSDALE	NY	10530	NOT EMPLOYED
4 IND	KOLBE, NINA	WASHINGTON	DC	20003	SELF
5 IND	ESSENBERG, PENNY	GRAND RAPIDS	MI	49534	
6 IND	GIEBEL, BETTY	KISSIMMEE	FL	34744	HOMEMAKER
OCCUPATION		TRANSACTION_DT	TRANSACTION_AMT	OTHER_ID	TRAN_ID
1 NOT EMPLOYED		7112019	15	C00042366	SA11AI_165851302
2 NOT EMPLOYED		3042020	39	C00696948	SA11AI_216676150
3 NOT EMPLOYED		12182019	15		6260719
4 HEALTH PROFESSIONAL		8302019	10	C00637074	SA11AI_174544101
5		3212019	16	C00580068	SB28A_145393165
6 HOMEMAKER		7112019	42		SA11AI.178611
FILE_NUM	MEMO_CD				
1 1378435					
2 1402724					
3 1371576					
4 1378435					
5 1344765					
6 1358506					
MEMO_TEXT					
1 EARMARKED FOR DSCC (C00042366)					
2 EARMARKED FOR BERNIE 2020 (C00696948)					
3					
4 EARMARKED FOR FINKENAUER FOR CONGRESS (C00637074)					
5 REFUND OF CONTRIBUTION, INITIALLY EARMARKED FOR PROGRESSIVE TURNOUT PROJECT (C00580068)					
6					
SUB_ID	newdates	citystatepair			
1 4.03022e+18	2019-07-11	CAPE CORAL, FL			
2 4.05142e+18	2020-03-04	SCOTTSVILLE, NY			
3 4.01222e+18	2019-12-18	HARTSDALE, NY			
4 4.03022e+18	2019-08-30	WASHINGTON, DC			
5 4.08282e+18	2019-03-21	GRAND RAPIDS, MI			
6 4.10182e+18	2019-07-11	KISSIMMEE, FL			

In [51]: `w <- tapply(myDF$TRANSACTION_AMT, myDF$citystatepair, sum)`In [53]: `head(w)`

# B, CA	\$0.00, PA	\$20.00, PA
200	122	40
\$49.02, PA	\$72.53, PA	'CRESCENT CITY' CA, CA
98	145	81

In [54]: `head(w, n=50)`

# B, CA	\$0.00, PA
200	122
\$20.00, PA	\$49.02, PA
40	98
\$72.53, PA	'CRESCENT CITY' CA, CA
145	81
(TACOMA) UNIVERSIT, WA	*CITY, IN
20	50
,	, AZ
287142	300
, BEN LOMONDX, CA	, CA
217	12485
, DC	, FL
3200	37300
, HI	, IL
500	230
, MD	, MN
500	250
, NJ	, NY
250	1250
, PA	, TX
326	200
, VA	, KEOKUK, IA
600	60
--SELECT--, OH	-ST AUGUSTINE, FL
270	15
.BRISTOL VT, VT	.CAMERON, WI
95	150
. TAMPA, FL	000000L, WI
50	169
00184 ROME 0 ITALY, ZZ	00184 ROME, ZZ
500	750
007 JAPAN, ZZ	01-07 DAYS, OR
550	81
OVERLAND PARK, KS	1 ELK AVENUE, NY
510	250
10025-4114 USA, NY	10C, NY
72	100
117570 SINGAPORE,	123 E 156TH STREET, OH
2500	25
1296 COPPET AA 0 SWITZERLAND, ZZ	1296 COPPET, AA
100	100
13626 SKY HAWK DR, AZ	14 A, NY
1	300
15705 SANTIAGO DE , MD	158 AUSTRALIA, ZZ
50	15
15A, NY	16 HAWKVIEW ROAD, NH
147	250
18B, NY	1900 W GATESBURG R, PA
25	215

In [60]: `head(w, n=20)`

project\_5

# B, CA	\$0.00, PA	\$20.00, PA
200	122	40
\$49.02, PA	\$72.53, PA	'CRESCENT CITY' CA, CA
98	145	81
(TACOMA) UNIVERSIT, WA	*CITY, IN	,
20	50	287142
, AZ	, BEN LOMONDX, CA	, CA
300	217	12485
, DC	, FL	, HI
3200	37300	500
, IL	, MD	, MN
230	500	250
, NJ	, NY	
250	1250	

In [55]: `tail(w, n=50)`

ZIGZAG, OR	ZILLAH, WA
10	2173
ZIM, MN	ZIMMERMAN, MN
10794	11176
ZINI, NM	ZION CROSSROADS, VA
150	1437
ZION GROVE, PA	ZION, IL
31	5016
ZIONISVILLE, IN	ZIONSVILE, IN
1000	25
ZIONSVILLE, IN	ZIONSVILLE, PA
173224	2865
ZIONSVILLE, IN	ZIONVILLE, IN
1519	293
ZIONVILLE, NC	ZIONVILLE, ZZ
956	10
ZIRCONIA, NC	ZLOS ANGELES, CA
2297	145
ZNEWTON, NC	ZOAR, OH
10	1000
ZOLFO SPRINGS, FL	ZONSVILLE, IN
440	35
ZORTMAN, MT	ZUERICH, MD
92	15
ZUERICH, ZZ	ZUG 60618 SWITZERLAND, ZZ
3	118
ZUMBRO FALLS, MN	ZUMBROTA, MN
461	1133
ZUNI, NM	ZUNI, VA
3451	300
ZURICH 8005 SWITZERLAND, ZZ	ZURICH 8006 SWITZERLAND, ZZ
27	100
ZURICH 8008 SWITZERLAND, ZZ	ZURICH 8032 SWITZERLAND, ZZ
100	125
ZURICH 8038 SWITZERLAND, ZZ	ZURICH 8049 SWITZERLAND, ZZ
2800	1500
ZURICH 8051 SWITZERLAND, ZZ	ZURICH 8052 SWITZERLAND, ZZ
177	27
ZURICH CH-8032 SWITZERLAND, ZZ	ZURICH SWITZERLAND, ZZ
100	35
ZURICH, KS	ZURICH, ZZ
300	5608
ZWINGLE, IA	ZWOLLE, LA
963	3346
`ALBANY, CA	`NEW YORK, NY
37	2800
`OAK PARK, IL	`SEMINOLE, FL
15	131
{ACIFIC PALISADES, CA	{ASADENA, CA
100	500

In [61]: `tail(w, n=20)`

ZURICH 8005 SWITZERLAND, ZZ	27	ZURICH 8006 SWITZERLAND, ZZ	100
ZURICH 8008 SWITZERLAND, ZZ	100	ZURICH 8032 SWITZERLAND, ZZ	125
ZURICH 8038 SWITZERLAND, ZZ	2800	ZURICH 8049 SWITZERLAND, ZZ	1500
ZURICH 8051 SWITZERLAND, ZZ	177	ZURICH 8052 SWITZERLAND, ZZ	27
ZURICH CH-8032 SWITZERLAND, ZZ	100	ZURICH SWITZERLAND, ZZ	35
ZURICH, KS	300	ZURICH, ZZ	5608
ZWINGLE, IA	963	ZWOLLE, LA	3346
`ALBANY, CA	37	`NEW YORK, NY	2800
`OAK PARK, IL	15	`SEMINOLE, FL	131
{ACIFIC PALISADES, CA	100	{ASADENA, CA	500

In [57]: `head(sort(w, decreasing=TRUE))`

NEW YORK, NY SAN FRANCISCO, CA	331576505	WASHINGTON, DC	LOS ANGELES, CA
	38701698	37152618	14141160
HOUSTON, TX	DALLAS, TX		
12463126	11642460		

In [58]: `head(sort(w, decreasing=TRUE), n=20)`

NEW YORK, NY SAN FRANCISCO, CA	331576505	WASHINGTON, DC	LOS ANGELES, CA
	38701698	37152618	14141160
HOUSTON, TX	DALLAS, TX	CHICAGO, IL	BOSTON, MA
12463126	11642460	11032607	8287423
ATLANTA, GA	SEATTLE, WA	SAN ANTONIO, TX	LAS VEGAS, NV
7141597	7115650	6467233	5699480
DENVER, CO	AUSTIN, TX	BROOKLYN, NY	ARLINGTON, VA
5643729	5555237	5416132	5217976
BALA CYNWYD, PA	PALO ALTO, CA	PHILADELPHIA, PA	GREENWICH, CT
5143506	4975339	4357885	3942437

In [59]: `tail(sort(w, decreasing=TRUE), n=20)`

EXETER, PA	EASTER, FL	COOL, CA	CORTEX, FL
-1357	-1500	-1681	-1800
SUMPTER TWP, MI	FREEMANSBURG, PA	WELLBORN, FL	CAMERON, TX
-1850	-2003	-2056	-2157
WILMINTON, NC	GORDON, TX	LOOKOUT MTN, GA	E. AMHERST, NY
-2300	-2338	-2373	-2700
BLVD, OR	BROOKVILLE, KS	DARIEN, NY	GARDEN CITY, NJ
-2800	-2800	-2800	-2800
WILSON, AR	PLEASANT HILL, TN	HOXIE, KS	HEFLIN, AL
-2800	-4295	-8150	-9983

Find the sum of the total donations contributed in each state.

Create a new column that pastes together the city and state.

```
[1] "CAPE CORAL, FL" "SCOTTSVILLE, NY" "HARTSDALE, NY" "WASHINGTON, DC"
[5] "GRAND RAPIDS, MI" "KISSIMMEE, FL"
```

Find the total donation amount for each city/state location. In the output do you notice anything suspicious in the result? From the first 20, there are some missing data. after sorting, looking at the tail some are negative. How do you think that occurred? due to omission and has some have some money to be returned to them

## Question 4

In [62]: `head(myDF$NAME)`

```
[1] "SOLOMON, VICTOR"  "MYERS, ELIZABETH" "HASKINS, GEORGE"  "KOLBE, NINA"
[5] "ESSENBERG, PENNY" "GIEBEL, BETTY"
```

In [63]: `head(myDF$NAME, n=50)`

[1]	"SOLOMON, VICTOR"	"MYERS, ELIZABETH"
[3]	"HASKINS, GEORGE"	"KOLBE, NINA"
[5]	"ESSENBERG, PENNY"	"GIEBEL, BETTY"
[7]	"GRISMORE, STEVE"	"POCIECHA-PALM, CAROL"
[9]	"DICE, KEVIN"	"KULSTAD, ERIK"
[11]	"SMELLEY, KAY MS."	"WELLMAN, NANCY"
[13]	"WARD, KATHLEEN A. MRS."	"GIBLIN, KATHLEEN"
[15]	"KELLEY, PATRICK RUSTON MR."	"WILSON, RICHARD J"
[17]	"DRUMMOND, F FORD MR."	"KROKEN, DUANE"
[19]	"BIRD, KENNETH"	"FALCHUK, EVAN"
[21]	"MESELE, SAMSON"	"HAWKINS, FRANK"
[23]	"KNOX, RICHARD G. MR."	"MOORE, JOAN"
[25]	"MCNULTY, KELLY"	"HUDSON, PEGGY M"
[27]	"ASMUS, GINGER"	"ELLINGSEN, PAUL"
[29]	"BRENEMAN, JAY"	"LEVEN, SEYMOUR"
[31]	"LUTZ, CHRISTOPHER"	"MILLER, VALERIE"
[33]	"JEANRENAUD, BARBARA"	"REILLY, JOE"
[35]	"TALBOTT, JOHN"	"FINK, MICHAEL"
[37]	"LULOW, ROBERT"	"HORNNSBY, RICHARD"
[39]	"MORROW, CLARENCE"	"LOVE, WILLIE"
[41]	"FLOREA, IOAN"	"PATTERSON, SASHA"
[43]	"VOLANTE, MONICA"	"WYSCAVER, NIKKI"
[45]	"PIRONKOVA IRWIN, MARIA"	"HOLMES, SANDY"
[47]	"NELSON, BONNIE A"	"CHACE, DONALD"
[49]	"MARTIN, DON S"	"EMMETT, CONSTANCE"

In [64]: `class(myDF$NAME)`

```
[1] "character"
```

In [67]: `grep("DON", head(myDF$NAME, n=50))`

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

In [68]: `sum(grep("DON", head(myDF$NAME, n=50)))`

```
[1] 2
```

In [71]: `sum(grep("DON", head(myDF$NAME, n=1000)))`

```
[1] 7
```

```
In [73]: sum(grepl(", MARY", head(myDF$NAME, n=1000)))
# you can put any name you want.
```

```
[1] 12
```

```
In [77]: sum(grepl(", MARY", myDF$NAME))
# taking the head off and searching the entire column
```

```
[1] 57516
```

```
In [78]: sum(grepl(", MARY ", myDF$NAME))
# you can play around by putting space after the name mary
# but the values will change, it depends on how you want to do it.
```

```
[1] 18031
```

```
In [79]: head(grepl(", MARY ", myDF$NAME))
```

```
[1] FALSE FALSE FALSE FALSE FALSE FALSE
```

```
In [82]: sum(myDF$TRANSACTION_AMT[grepl(", MARY ", myDF$NAME)])
# it means if i search for ", MARY " for which is TRUE,
# i will go and extract the TRANSACTION_AMT
```

```
[1] 4007115
```

```
In [83]: length(myDF$TRANSACTION_AMT[grepl(", MARY ", myDF$NAME)])
# to compare with value gotten before.
```

```
[1] 18031
```

Find the type of data that is in the NAME column.

```
[1] "character"
```

Split up the names in the NAME column, to extract the first names of the donors. (This will not be perfect, but it is our first attempt.)

How much money is donated (altogether) by people named Mary? [1] 4007115

## Question 5

```
In [84]: head(myDF)
```

CMTE_ID	AMNDT_IND	RPT_TP	TRANSACTION_PGI	IMAGE_NUM	TRANSACTION_TP
1 C00401224	N	YE		2.020013e+17	24T
2 C00401224	N	M4	P	2.020042e+17	24T
3 C00193433	N	YE	P	2.020012e+17	15
4 C00401224	N	YE		2.020013e+17	24T
5 C00401224	N	MY		2.019073e+17	24T
6 C00618371	N	Q3	P	2.019102e+17	15
	ENTITY_TP	NAME	CITY	STATE	ZIP_CODE EMPLOYER
1 IND	SOLOMON, VICTOR	CAPE CORAL	FL	33904	RETIRED
2 IND	MYERS, ELIZABETH	SCOTTSVILLE	NY	14546	NOT EMPLOYED
3 IND	HASKINS, GEORGE	HARTSDALE	NY	10530	NOT EMPLOYED
4 IND	KOLBE, NINA	WASHINGTON	DC	20003	SELF
5 IND	ESSENBERG, PENNY	GRAND RAPIDS	MI	49534	
6 IND	GIEBEL, BETTY	KISSIMMEE	FL	34744	HOMEMAKER
	OCCUPATION		TRANSACTION_DT	TRANSACTION_AMT	OTHER_ID TRAN_ID
1 NOT EMPLOYED		7112019	15		C00042366 SA11AI_165851302
2 NOT EMPLOYED		3042020	39		C00696948 SA11AI_216676150
3 NOT EMPLOYED		12182019	15		6260719
4 HEALTH PROFESSIONAL		8302019	10		C00637074 SA11AI_174544101
5		3212019	16		C00580068 SB28A_145393165
6 HOMEMAKER		7112019	42		SA11AI.178611
	FILE_NUM	MEMO_CD			
1 1378435					
2 1402724					
3 1371576					
4 1378435					
5 1344765					
6 1358506					
	MEMO_TEXT				
1 EARMARKED FOR DSCC (C00042366)					
2 EARMARKED FOR BERNIE 2020 (C00696948)					
3					
4 EARMARKED FOR FINKENAUER FOR CONGRESS (C00637074)					
5 REFUND OF CONTRIBUTION, INITIALLY EARMARKED FOR PROGRESSIVE TURNOUT PROJECT (C00580068)					
6					
	SUB_ID	newdates	citystatepair		
1 4.03022e+18	2019-07-11	CAPE CORAL, FL			
2 4.05142e+18	2020-03-04	SCOTTSVILLE, NY			
3 4.01222e+18	2019-12-18	HARTSDALE, NY			
4 4.03022e+18	2019-08-30	WASHINGTON, DC			
5 4.08282e+18	2019-03-21	GRAND RAPIDS, MI			
6 4.10182e+18	2019-07-11	KISSIMMEE, FL			

In [85]: `head(myDF$EMPLOYER)`

```
[1] "RETIRED"      "NOT EMPLOYED" "NOT EMPLOYED" "SELF"          ""
[6] "HOMEMAKER"
```

In [86]: `head(myDF$EMPLOYER, n= 50)`

```
[1] "RETIRED"
[2] "NOT EMPLOYED"
[3] "NOT EMPLOYED"
[4] "SELF"
[5] ""
[6] "HOMEMAKER"
[7] "U OF IOWA"
[8] "STATE OF WISCONSIN DEPARTMENT OF CORRE"
[9] "IRON MOUNTAIN"
[10] "ATTUNE"
[11] "AMERICAN ALLOY STEEL"
[12] "SELF"
[13] "SELF-EMPLOYED"
[14] "NONE"
[15] "AMERICAN CHEMISTRY COUNCIL"
[16] "HONEYWELL INTERNATIONAL"
[17] "BANCFIRST"
[18] "RETIRED"
[19] "US GOVT"
[20] "VILLAGEPLAN"
[21] "WACHTELL, LIPTON, ROSEN & KATZ"
[22] "SCALAR MEDIA PARTNERS"
[23] "RETIRED"
[24] "N/A"
[25] "EVOKE NAVIENCE"
[26] "SALEM RADIO NETWORK"
[27] "RETIRED"
[28] "DAD'S PLACE"
[29] "STRATEGY SOLUTIONS INC"
[30] "RETIRED"
[31] ""
[32] "BOSTONS CHILDRENS HOSPITAL"
[33] "RETIRED"
[34] "WORLDWIDE FLIGHT SERVICES"
[35] "INDIANA UNIVERSITY"
[36] ""
[37] "RETIRED"
[38] "PRIVATE INFORMATION"
[39] "RETIRED"
[40] "NOT EMPLOYED"
[41] "SELF-EMPLOYED"
[42] "NOT EMPLOYED"
[43] "PHILIPS NORTH AMERICA"
[44] "NOT EMPLOYED"
[45] "INMARSAT"
[46] "DECKER TRUCK LINE"
[47] ""
[48] "MEDOLAC LABORATORIES"
[49] "AMFM BROADCASTING INC."
[50] "NOT EMPLOYED"
```

In [88]: `sum(myDF$TRANSACTION_AMT[myDF$EMPLOYER == "NOT EMPLOYED"])`

```
[1] NA
```

In [89]: `sum(myDF$TRANSACTION_AMT[myDF$EMPLOYER == "NOT EMPLOYED"], na.rm=TRUE)`  
*# to remove NA*

```
[1] 79250204
```

```
In [92]: sum(myDF$TRANSACTION_AMT[myDF$EMPLOYER == "EMPLOYED"], na.rm=TRUE)
# only those that wrote down that they are employed.
```

```
[1] 12418
```

```
In [93]: sum(myDF$TRANSACTION_AMT[myDF$EMPLOYER != "NOT EMPLOYED"], na.rm=TRUE)
# this value removed only those that are not
# employed and leave the rest of the people
# it is not perfect.
```

```
[1] 988014293
```

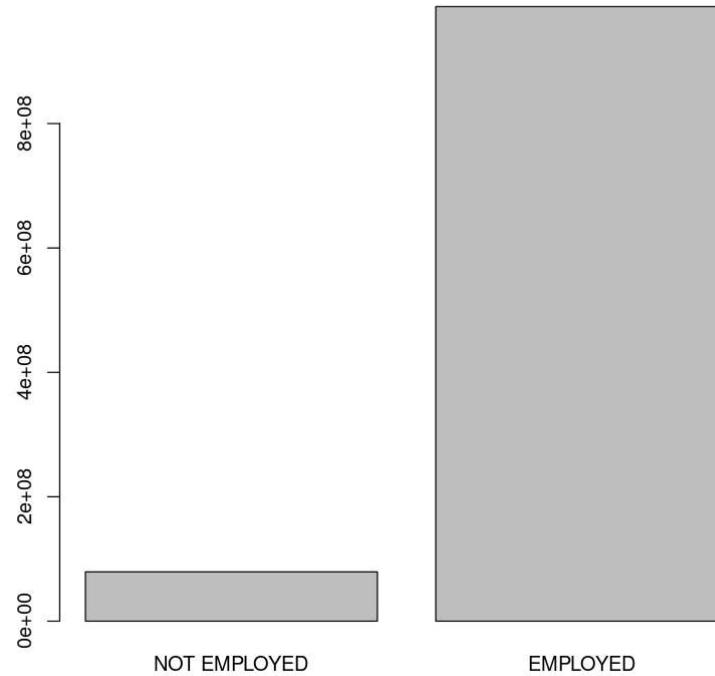
```
In [103... notemployed <- sum(myDF$TRANSACTION_AMT[myDF$EMPLOYER == "NOT EMPLOYED"],
na.rm=TRUE)
```

```
In [104... employed <- sum(myDF$TRANSACTION_AMT[myDF$EMPLOYER != "NOT EMPLOYED"],
na.rm=TRUE)
```

```
In [105... v <- c(notemployed, employed )
```

```
In [106... names(v) <- c("NOT EMPLOYED", "EMPLOYED")
```

```
In [107... barplot(v)
```



```
In [108... head(tapply(myDF$TRANSACTION_AMT, myDF$OCCUPATION, sum))
```

"COMPUTER TUTOR" TO FAMILIES	
77364020	14
"NOT EMPLOYED	"NOT EMPLOYED"
25	83
\$ELF EMPLOYED	%%OCCUPATION%%
60	3641

In [109]: `tail(tapply(myDF$TRANSACTION_AMT, myDF$OCCUPATION, sum))`

{PROFESSOR	{RFESSPR	{ROFESSOR
54	269	100
{{OCCUPATION,DEFAULTTTO {{OCCUPATION,DEFAULTTTO=}}	250	~~BLANK
98		25

In [50]: `head(sort(tapply(myDF$TRANSACTION_AMT, myDF$OCCUPATION, sum), decreasing=TRUE))`

FOUNDER	RETIRED	NOT EMPLOYED	ATTORNEY	CEO
279442327	98774206	84855951	77364020	31326953
				25055647

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

```
[1] "NOT EMPLOYED"      "NOT EMPLOYED"      "NOT EMPLOYED"
[4] "HEALTH PROFESSIONAL"  ""               "HOMEMAKER"
```

In [74]: `head(myDF$OCCUPATION, n=100)`

```
[1] "NOT EMPLOYED"
[2] "NOT EMPLOYED"
[3] "NOT EMPLOYED"
[4] "HEALTH PROFESSIONAL"
[5] ""
[6] "HOMEMAKER"
[7] "TEACHER"
[8] "OCCUPATIONAL THERAPIST"
[9] "SENIOR VICE PRESIDENT"
[10] "PHYSICIAN"
[11] "ADMINISTRATIVE ASSISTANT"
[12] "CLASSICAL MUSIC CONSULTANT & MANAGER"
[13] "SELF-EMPLOYED"
[14] "NONE"
[15] "DIRECTOR FEDERAL AFFAIRS"
[16] "SR DEVELOPMENT PROGRAM MANAGER"
[17] "DIRECTOR"
[18] "RETIRED"
[19] "GENERAL ENGINEER"
[20] "CEO"
[21] "ATTORNEY"
[22] "CONSULTANT"
[23] "RETIRED"
[24] "RETIRED"
[25] "HEALTHCARE CONSULTANT"
[26] "VICE PRESIDENT OF NATIONAL SALES"
[27] "RETIRED"
[28] "RESTAURANT OWNER"
[29] "CONSULTANT"
[30] "PHYSICIAN"
[31] ""
[32] "RN"
[33] "RETIRED"
[34] "STATION MANAGER"
[35] "PROFESSOR"
[36] ""
[37] "RETIRED"
[38] "PROJECT MANAGER"
[39] "RETIRED"
[40] "NOT EMPLOYED"
[41] "HVAC"
[42] "NOT EMPLOYED"
[43] "DIRECTOR GOVERNMENT AFFAIRS"
[44] "NOT EMPLOYED"
[45] "ACCOUNTANT"
[46] "TRANSPORTATION"
[47] ""
[48] "CSO"
[49] "MARKET SVP PROGRAMMING"
[50] "NOT EMPLOYED"
[51] "NOT EMPLOYED"
[52] "NOT EMPLOYED"
[53] "CUSTOMER ADVISOR"
[54] "HORSE TRAINER"
[55] "CEROS FINANCIAL SERVICES"
[56] "AVP SYSTEM DEVELOPMENT"
[57] "PARTNER"
[58] "CONSTRUCTION"
[59] "ARCHITECT"
[60] "RETIRED"
```

```
[61] "HEALTHCARE INTERPRETER"
[62] "RETIRED"
[63] "RETIRED"
[64] "NOT EMPLOYED"
[65] ""
[66] "UNION SET PAINTER"
[67] "RETIRED"
[68] "NOT EMPLOYED"
[69] "ENVIRONMENTAL CONSULTANT"
[70] "NOT EMPLOYED"
[71] "LAWYER"
[72] "PSYCHOLOGIST"
[73] "MANAGER"
[74] "NOT EMPLOYED"
[75] "MT"
[76] ""
[77] "RETIRED"
[78] "STUDENT"
[79] "RETIRED"
[80] "NOT EMPLOYED"
[81] "NOT EMPLOYED"
[82] "PASTOR"
[83] "REAL ESTATE"
[84] "VISUAL MARKETING"
[85] "NOT EMPLOYED"
[86] "WEAR HOUSE WORKER"
[87] "NOT EMPLOYED"
[88] "LICENSEE"
[89] "ATTORNEY"
[90] "NOT EMPLOYED"
[91] "RETIRED"
[92] "SALES"
[93] "NONE"
[94] "PROFESSOR"
[95] "ARTIST"
[96] "CONSULTANT"
[97] "SELF EMPLOYED"
[98] "RETIRED"
[99] "NOT EMPLOYED"
[100] "CONSULTANT"
```

In [78]: `sum(myDF$TRANSACTION_AMT[myDF$OCCUPATION == "ACCOUNTANT"])`

```
[1] NA
```

In [79]: `sum(myDF$TRANSACTION_AMT[myDF$OCCUPATION == "ACCOUNTANT"], na.rm=TRUE)`

```
[1] 1067554
```

In [80]: `sum(myDF$TRANSACTION_AMT[myDF$OCCUPATION == "PROFESSOR"], na.rm=TRUE)`

```
[1] 4808341
```

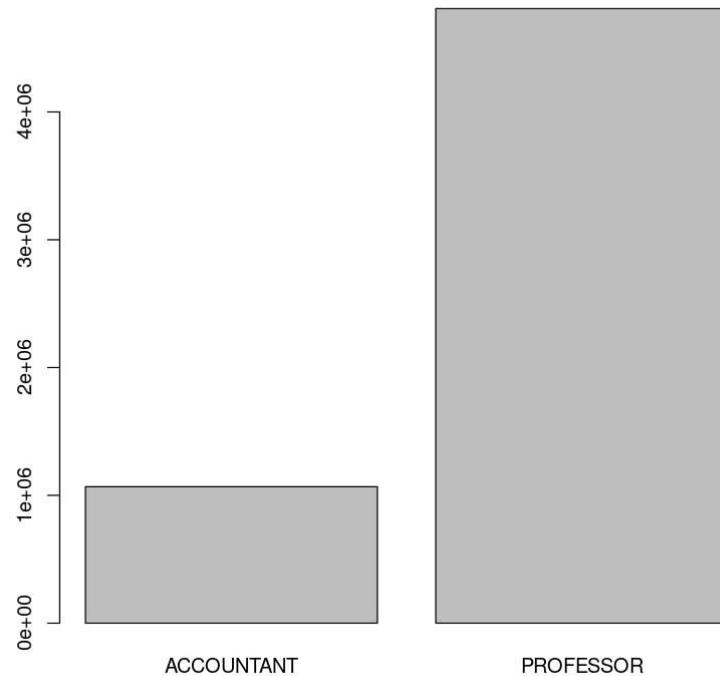
In [81]: `accountant <- sum(myDF$TRANSACTION_AMT[myDF$OCCUPATION == "ACCOUNTANT"], na.rm=TRUE)`

In [82]: `professor <- sum(myDF$TRANSACTION_AMT[myDF$OCCUPATION == "PROFESSOR"], na.rm=TRUE)`

```
In [83]: z <- c(accountant, professor )
```

```
In [84]: names(z) <- c("ACCOUNTANT", "PROFESSOR")
```

```
In [85]: barplot(z)
```



Using a barplot or dotchart, show the total amount of donations made by EMPLOYED vs NOT EMPLOYED individuals

What is the category of occupation that donates the most money?  
FOUNDER with 279442327

Plot something that you find interesting about the employment and/or occupation columns. I used bar plot to show the total amount of donations made by ACCOUNTANT vs PROFESSOR

## 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.