

Group: True Detectives

Initial scraped data

	CASENO	OFFENSE	EVENTDT	EVENTTT	CVLEGEND	CVDOW	InDbDate	Block_Location	BLKADDR
1	16017938	BURGLARY COMMERCIAL	03/28/2016 12:00:00 AM	10:15	BURGLARY - COMMERCIAL	1	05/01/2016 03:30:13 AM	2500 SHATTUCK AVE Berkeley, CA (37.863811, -122....	2500 SHATTUCK
2	16008050	BURGLARY RESIDENTIAL	02/09/2016 12:00:00 AM	16:27	BURGLARY - RESIDENTIAL	2	05/01/2016 03:30:11 AM	2100 HASTE ST Berkeley, CA (37.864908, -122.2672....	2100 HASTE ST
3	16024984	VANDALISM	04/27/2016 12:00:00 AM	18:00	VANDALISM	3	05/01/2016 03:30:08 AM	1700 BLAKE ST Berkeley, CA (37.861929, -122.276345)	1700 BLAKE ST
4	16005403	NARCOTICS	01/28/2016 12:00:00 AM	02:10	DRUG VIOLATION	4	05/01/2016 03:30:10 AM	1200 HASKELL ST Berkeley, CA (37.850541, -122.28....	1200 HASKELL S
5	16016426	VEHICLE STOLEN	03/20/2016 12:00:00 AM	12:59	MOTOR VEHICLE THEFT	0	05/01/2016 03:30:12 AM	2600 MILVIA ST Berkeley, CA (37.861843, -122.2696....	2600 MILVIA ST
6	16004530	THEFT MISD. (UNDER \$950)	01/22/2016 12:00:00 AM	21:59	LARCENY	5	05/01/2016 03:30:10 AM	1400 SHATTUCK AVE Berkeley, CA (37.881957, -122....	1400 SHATTUCK
7	16016313	NARCOTICS	03/19/2016 12:00:00 AM	23:30	DRUG VIOLATION	6	05/01/2016 03:30:12 AM	CEDAR STREET & BONITA AVE Berkeley, CA (37.8....	CEDAR STREET &
8	16017295	THEFT MISD. (UNDER \$950)	03/24/2016 12:00:00 AM	17:55	LARCENY	4	05/01/2016 03:30:12 AM	2000 OREGON ST Berkeley, CA (37.857409, -122.268....	2000 OREGON S
9	16010144	SEXUAL ASSAULT MISD.	02/20/2016 12:00:00 AM	02:25	SEX CRIME	6	05/01/2016 03:30:11 AM	2400 PARKER ST Berkeley, CA (37.863072, -122.260....	2400 PARKER ST
10	16000761	DISTURBANCE	01/05/2016 12:00:00 AM	06:58	DISORDERLY CONDUCT	2	05/01/2016 03:30:10 AM	2301 SYCAMORE ST Berkeley, CA	2301 SYCAMORE
11	15092183	BURGLARY AUTO	11/29/2015 12:00:00 AM	00:00	BURGLARY - VEHICLE	0	05/01/2016 03:30:08 AM	100 MAYBECK TWIN DR Berkeley, CA	100 MAYBECK T
12	15073156	BURGLARY RESIDENTIAL	12/15/2015 12:00:00 AM	00:00	BURGLARY - RESIDENTIAL	2	05/01/2016 03:30:09 AM	1600 BLAKE ST Berkeley, CA (37.861653, -122.278529)	1600 BLAKE ST
13	15092073	BURGLARY AUTO	11/09/2015 12:00:00 AM	03:00	BURGLARY - VEHICLE	1	05/01/2016 03:30:07 AM	BURNETT STREET & SAN PABLO AVE Berkeley, CA (37.861653, -122.278529)	BURNETT STREET
14	15065665	FRAUD/FORGERY	11/02/2015 12:00:00 AM	09:00	FRAUD	1	05/01/2016 03:30:07 AM	3100 SACRAMENTO ST Berkeley, CA (37.851477, -122....	3100 SACRAMENTO
15	16090617	BURGLARY AUTO	04/02/2016 12:00:00 AM	06:16	BURGLARY - VEHICLE	6	05/01/2016 03:30:13 AM	700 ALLSTON WAY Berkeley, CA (37.864703, -122.29....	700 ALLSTON W
16	16021155	DISTURBANCE	04/11/2016 12:00:00 AM	19:33	DISORDERLY CONDUCT	1	05/01/2016 03:30:13 AM	SHATTUCK AVENUE & CENTER ST Berkeley, CA (37.863072, -122.260....	SHATTUCK AVENUE

Challenges in Cleaning:

1. Extract "date" variable and store as "date" object
2. Store time as numeric object, converting minutes to decimals
3. Filter out cases without coordinates
4. Extract coordinates from "Block_location" variable
5. Separate latitude and longitude
6. Decide on severity for each crime
7. Decide on area groups for Berkeley (N, S, E, W)
8. Classify the location of each crime into one of the four quadrants

Data after cleaning

	date	day	time	type	details	lat	long	wday	severity	x	y	area
1	2016-03-28	Monday	10.25000000	BURGLARY - COMMERCIAL	BURGLARY COMMERCIAL	37.86381	-122.2674	weekday	2	0.005288	-0.007789	S
2	2016-02-09	Tuesday	16.45000000	BURGLARY - RESIDENTIAL	BURGLARY RESIDENTIAL	37.86491	-122.2673	weekday	4	0.005411	-0.006692	S
3	2016-04-27	Wednesday	18.00000000	VANDALISM	VANDALISM	37.86193	-122.2763	weekday	3	-0.003645	-0.009671	S
4	2016-01-28	Thursday	2.16666667	DRUG VIOLATION	NARCOTICS	37.85054	-122.2860	weekday	2	-0.013337	-0.021059	S
5	2016-03-20	Sunday	12.98333333	MOTOR VEHICLE THEFT	VEHICLE STOLEN	37.86184	-122.2696	weekend	4	0.003056	-0.009757	S
6	2016-01-22	Friday	21.98333333	LARCENY	THEFT MISD. (UNDER \$950)	37.88196	-122.2696	weekday	2	0.003149	0.010357	N
7	2016-03-19	Saturday	23.50000000	DRUG VIOLATION	NARCOTICS	37.87798	-122.2726	weekend	2	0.000142	0.006383	N
8	2016-03-24	Thursday	17.91666667	LARCENY	THEFT MISD. (UNDER \$950)	37.85741	-122.2685	weekday	2	0.004227	-0.014191	S
9	2016-02-20	Saturday	2.41666667	SEX CRIME	SEXUAL ASSAULT MISD.	37.86307	-122.2604	weekend	3	0.012348	-0.008528	E
10	2015-12-15	Tuesday	0.00000000	BURGLARY - RESIDENTIAL	BURGLARY RESIDENTIAL	37.86165	-122.2785	weekday	4	-0.005829	-0.009947	S
11	2015-11-09	Monday	3.00000000	BURGLARY - VEHICLE	BURGLARY AUTO	37.85301	-122.2869	weekday	3	-0.014177	-0.018592	S
12	2015-11-02	Monday	9.00000000	FRAUD	FRAUD/FORGERY	37.85148	-122.2787	weekday	1	-0.005951	-0.020123	S
13	2016-04-02	Saturday	6.26666667	BURGLARY - VEHICLE	BURGLARY AUTO	37.86470	-122.2997	weekend	3	-0.026954	-0.006897	W

Sample of code for cleaning (classifying the location of each crime into one of the four quadrants)

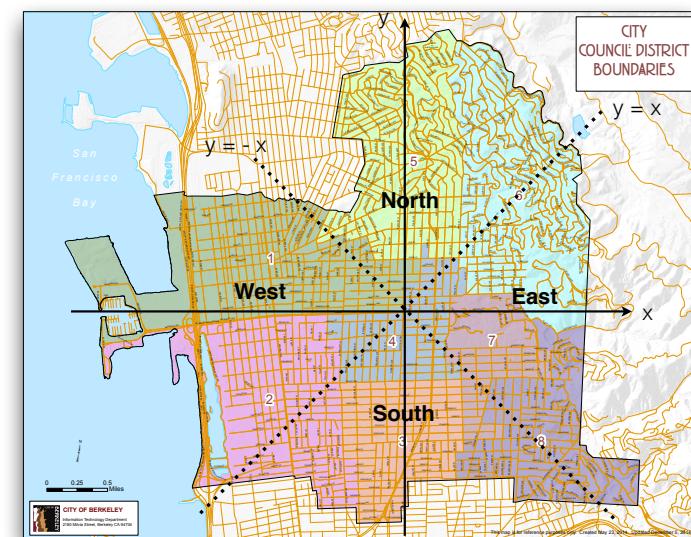
```
```{r}
Create new variables for x and y coordinates recentered by Berkeley's city centre
dc <- dc %>%
 mutate(x = long + 122.2727, y = lat - 37.8716)
Create new vectors to increase readability of code
x <- dc$x
y <- dc$y
area <- vector(length = length(x))

Categorising location into one of four quadrants
area[y >= 0 & -y <= x & x <= y] <- "N"
area[y < 0 & y <= x & x < -y] <- "S"
area[x < 0 & x < y & y <= -x] <- "W"
area[x >= 0 & -x <= y & y < x] <- "E"

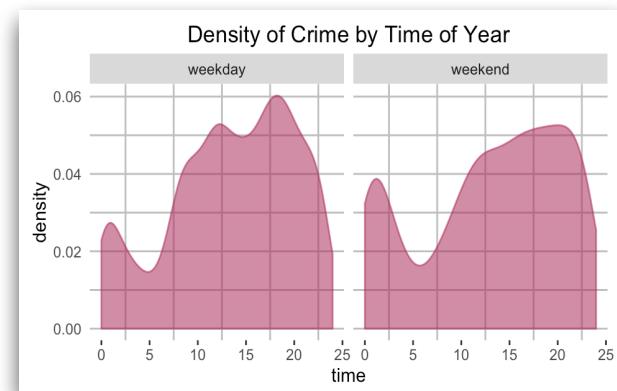
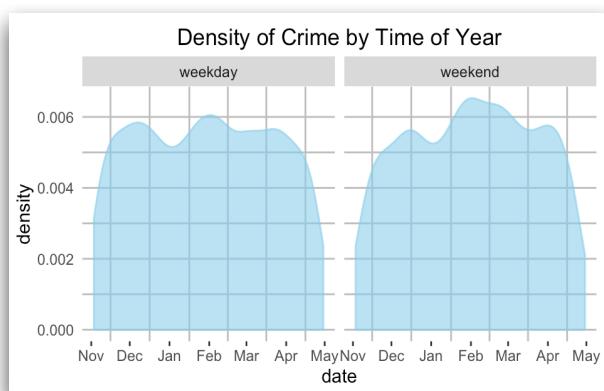
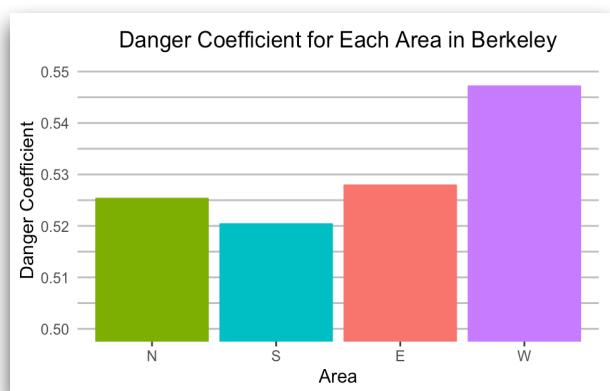
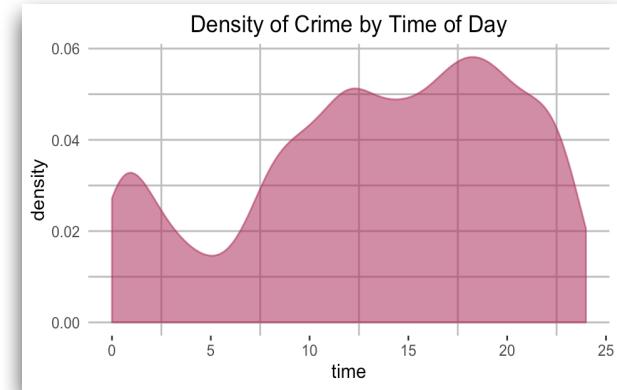
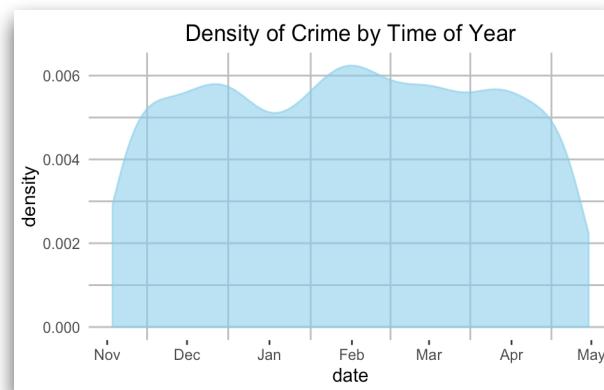
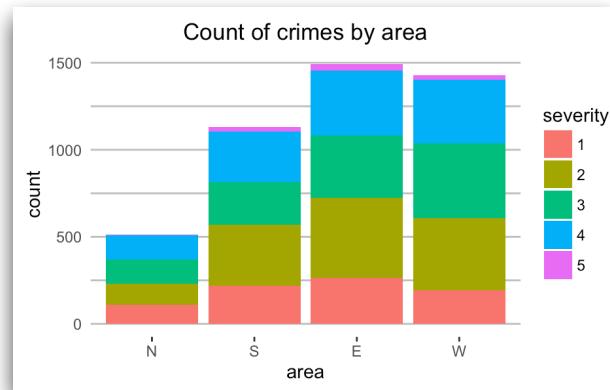
dc$area <- as.factor(area)
```

```

Map of Berkeley split into N, S, E, W



“What is the safest place to live in Berkeley? When is the safest time to roam around Berkeley?”



"If you wanted to commit a crime, where and when would be the perfect place to do it?"

