Northwestern Campus Safety Check

May 30th 2017

TEAM NEARBY

JASON DONG, WILL FINNEGAN, BEOMSEOK KIM, SAM KIM

Introduction: Objectives

We aim to give students a data-driven way to measure

the safety level around the campus

Introduction: Data Collection

Source of data: University Police Evanston Campus Blotter (http://www.northwestern.edu/up/blotter/blotter_ev.html)

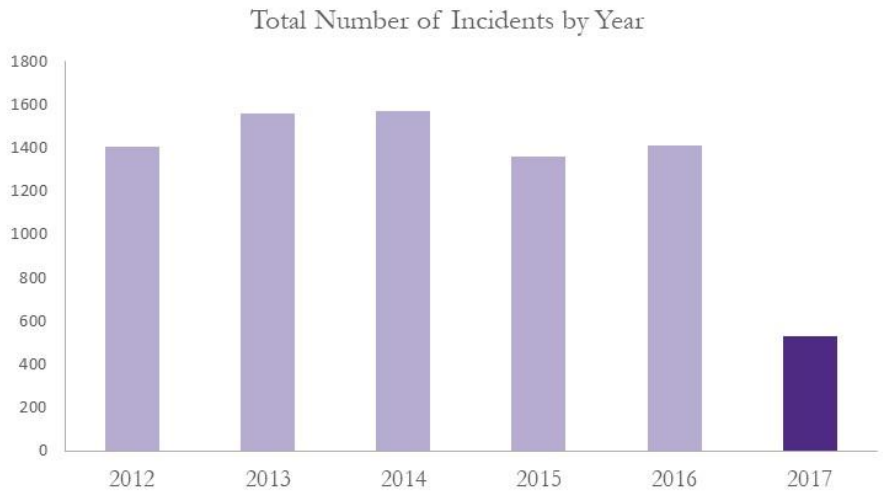
SECTION I: EVANSTON CAMPUS INCIDENTS

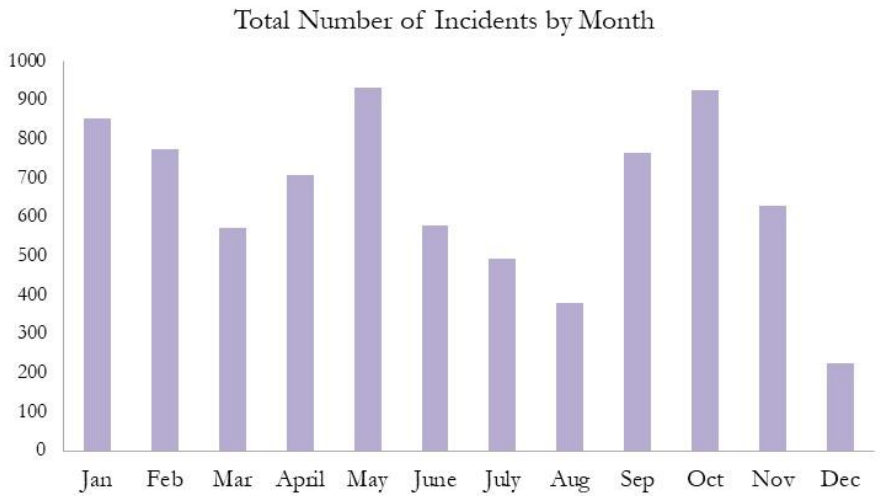
A daily log of incidents reported to the Evanston Police Department can be found at
http://www.cityofevanston.org/police/reports/daily-crime-reports/

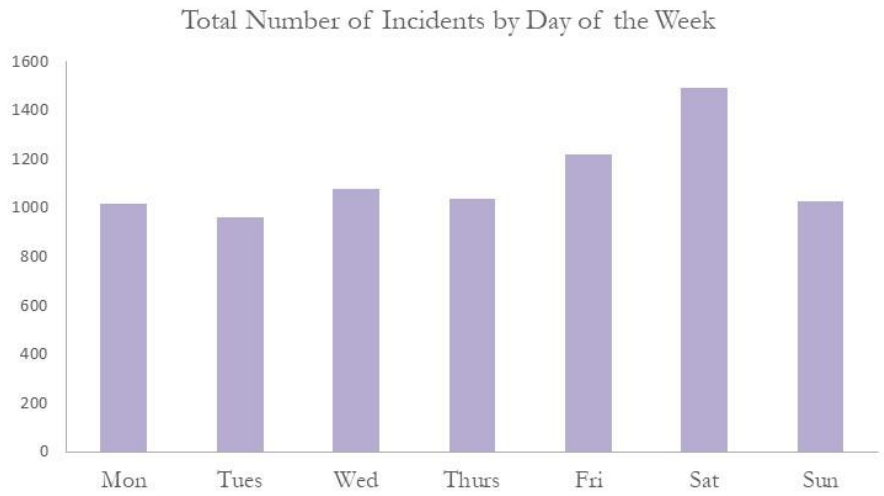
2017|apr|mar|feb|jan|
2016|dec|nov|oct|sep|aug|jul|jun|may|

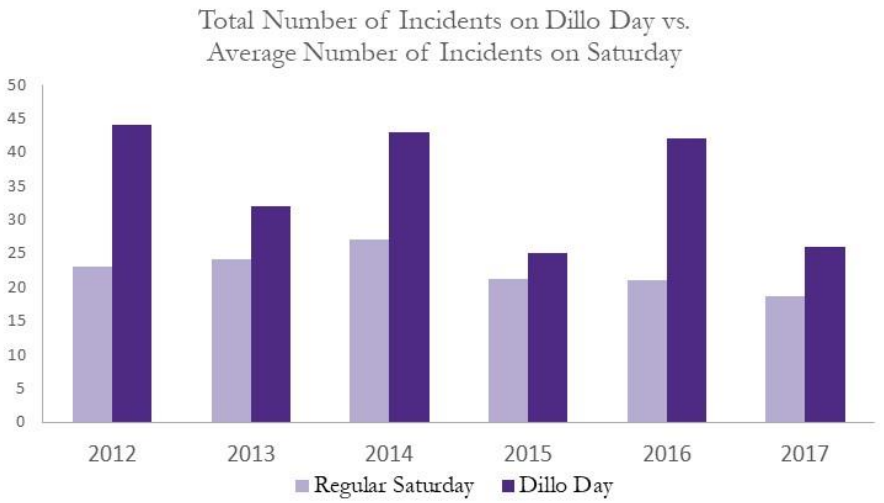
Case Number	**2017-00000613**
Date & Time: Reported	May 01, 2017 at 11:06:10 AM
Date & Time: Occurred	April 26, 2017 at 4:00:00 PM
	April 28, 2017 at 4:00:00 PM
Location:	2245 SHERIDAN RD
Common Name:	Sargent Hall
Incident Type:	Theft - Bicycle
Criminal Offense:	Theft $500 and Under
Disposition:	Closed

Brainstorming

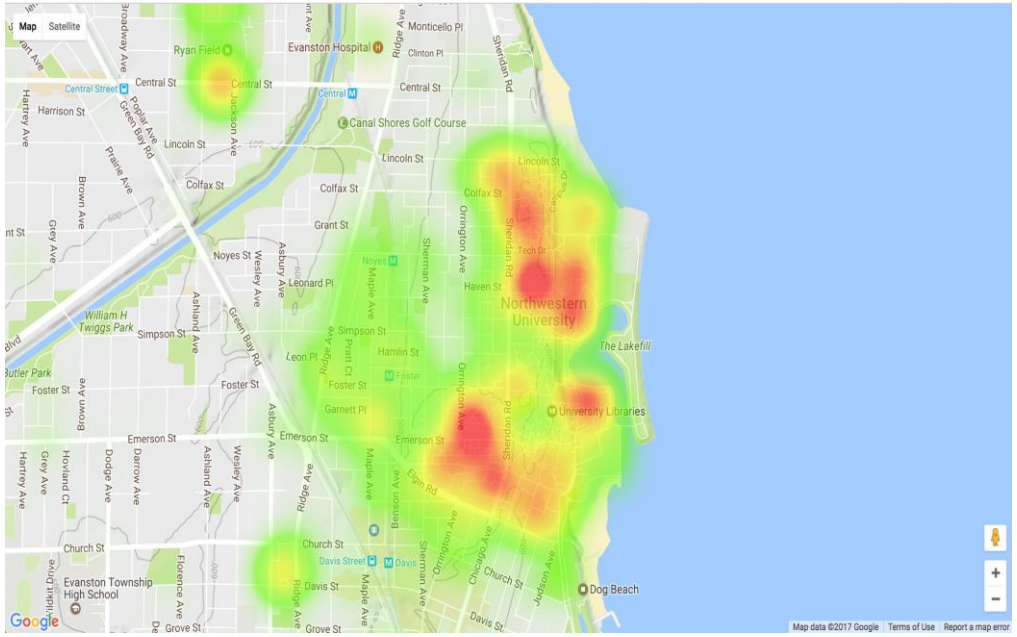




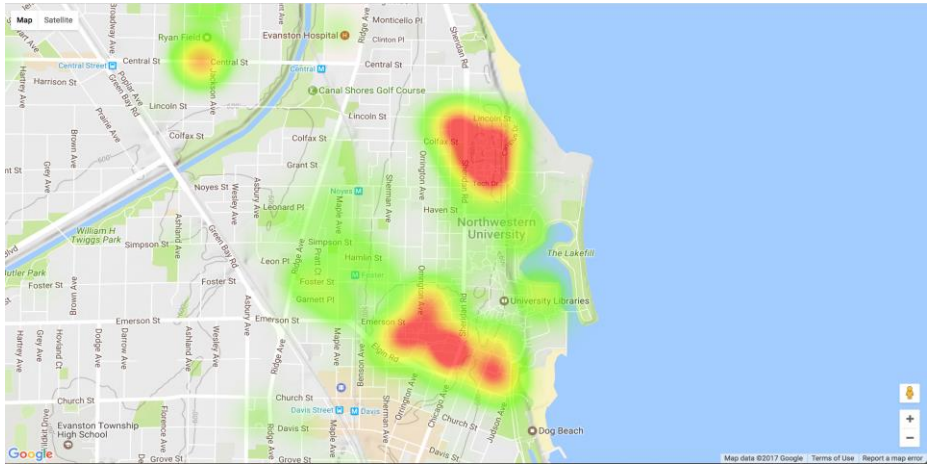




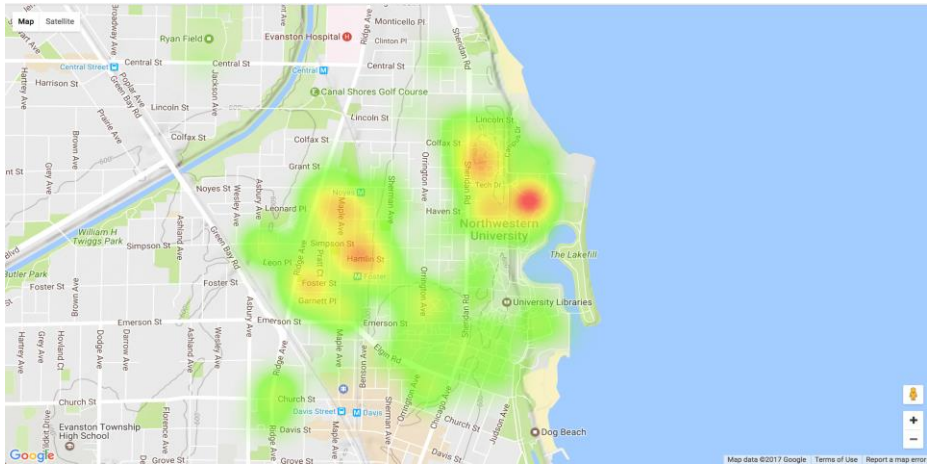
North Campus vs. South Campus



Heatmap of All Incidents



Heatmap of Substance-Related Incidents



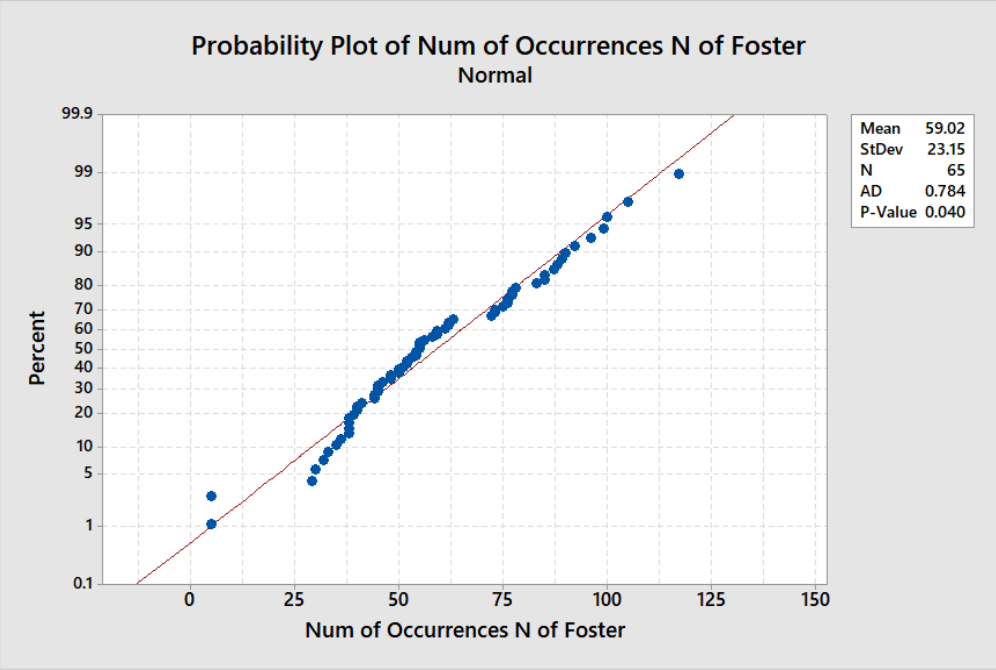
Heatmap of Noise-Related Incidents

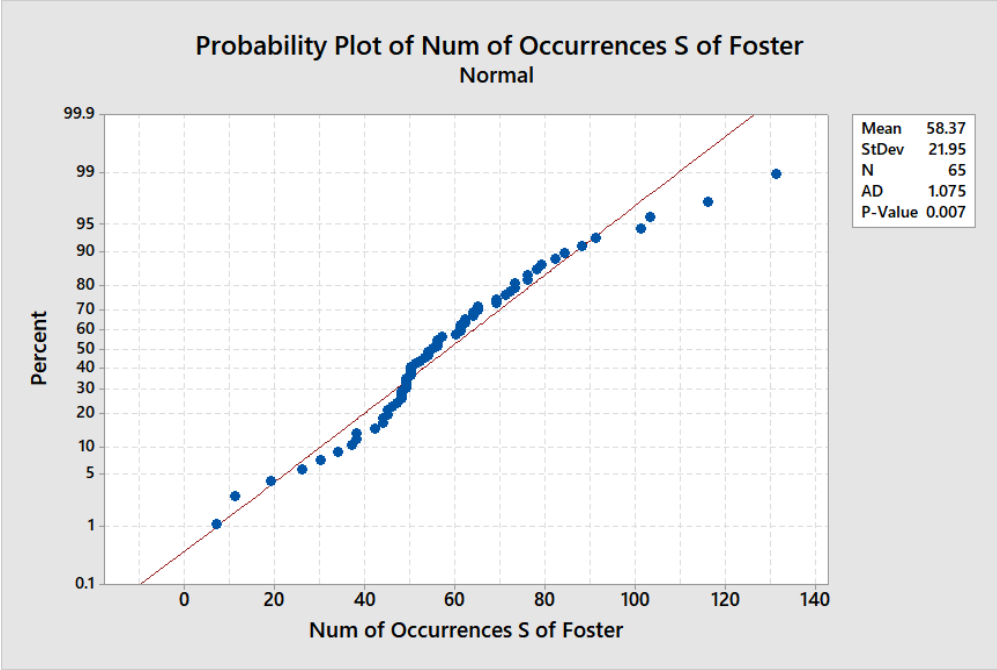
Sources: Google Maps.

Our Questions

- **Do more incidents take place on the North Campus compared to the South Campus?**
- **Do more incidents take place on the special occasions like Dillo Day? How do weekends compare to weekdays in general?**
- Do more incidents take place on the months when students are on break or months during classes?

North Campus vs. South Campus?





Approximately Normally Distributed

Sources: Minitab.

North Campus vs. South Campus?

Hypothesis Test

$H_0: \mu_1 = \mu_2$ *vs.* $H_1: \mu_1 \neq \mu_2$

(μ_1 : Average number of incidents North of Foster per month,
μ_2 : Average number of incidents South of Foster per month)

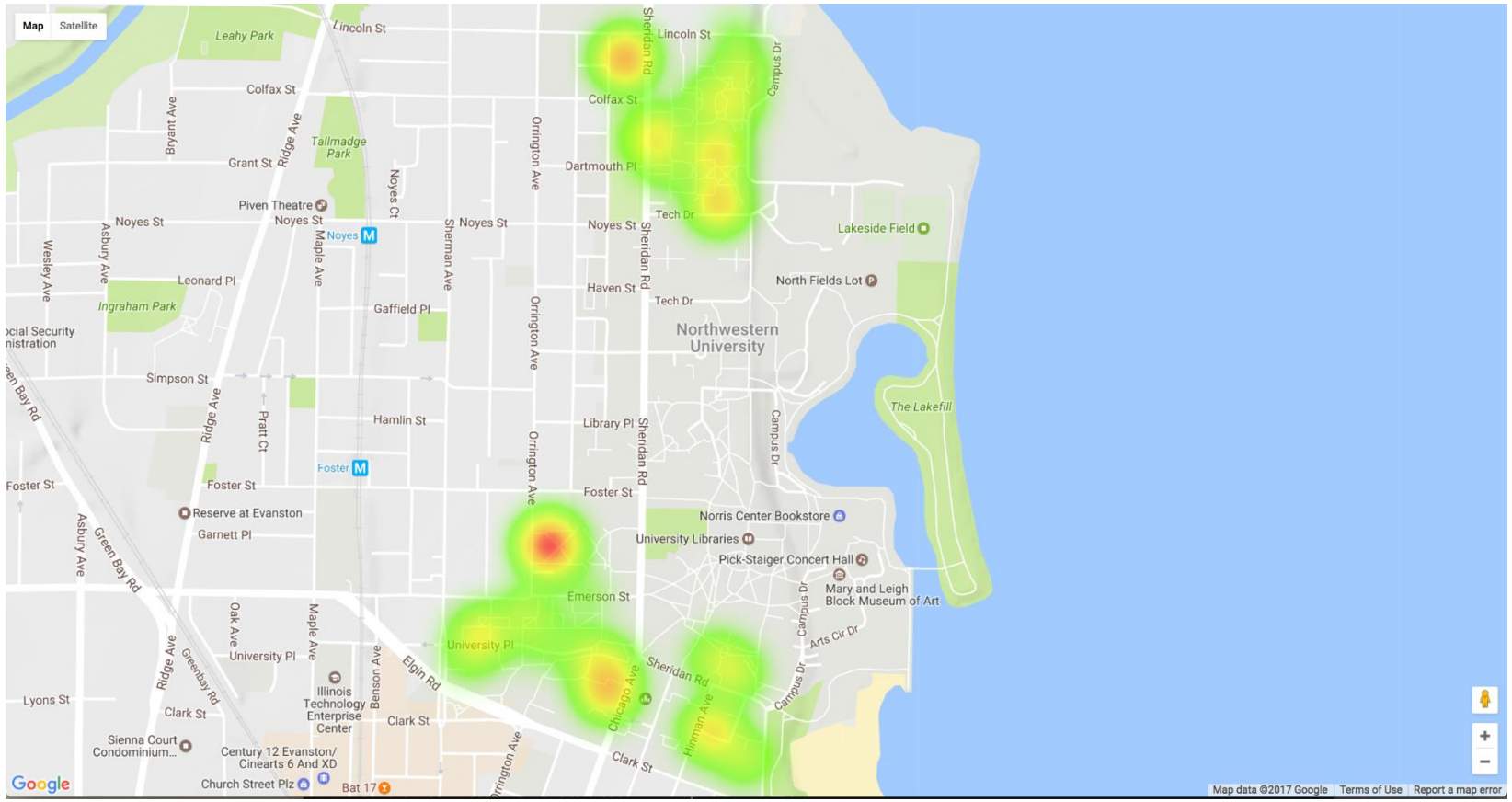
Two-Sample T-Test and CI: Num of Occurrences N of Foster, Num of Occurrences S of Foster

```
Two-sample T for Num of Occurrences N of Foster vs Num of Occurrences S of Foster

                              N  Mean  StDev  SE Mean
Num of Occurrences N of      65  59.0   23.1      2.9
Num of Occurrences S of      65  58.4   21.9      2.7


Difference = μ (Num of Occurrences N of Foster) - μ (Num of Occurrences S of Foster)
Estimate for difference:  0.65
95% CI for difference:  (-7.18, 8.47)
T-Test of difference = 0 (vs ≠): T-Value = 0.16  P-Value = 0.871  DF = 127
```

Fail to reject H_0 ; Statistically no significant difference

Sources: Minitab.

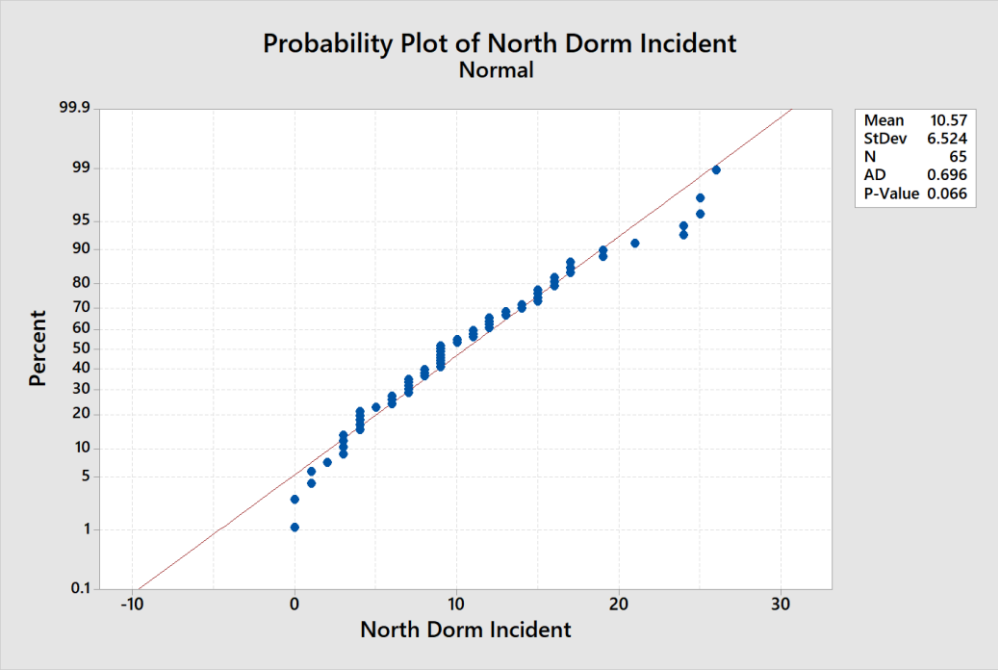
North Dorms vs. South Dorms?

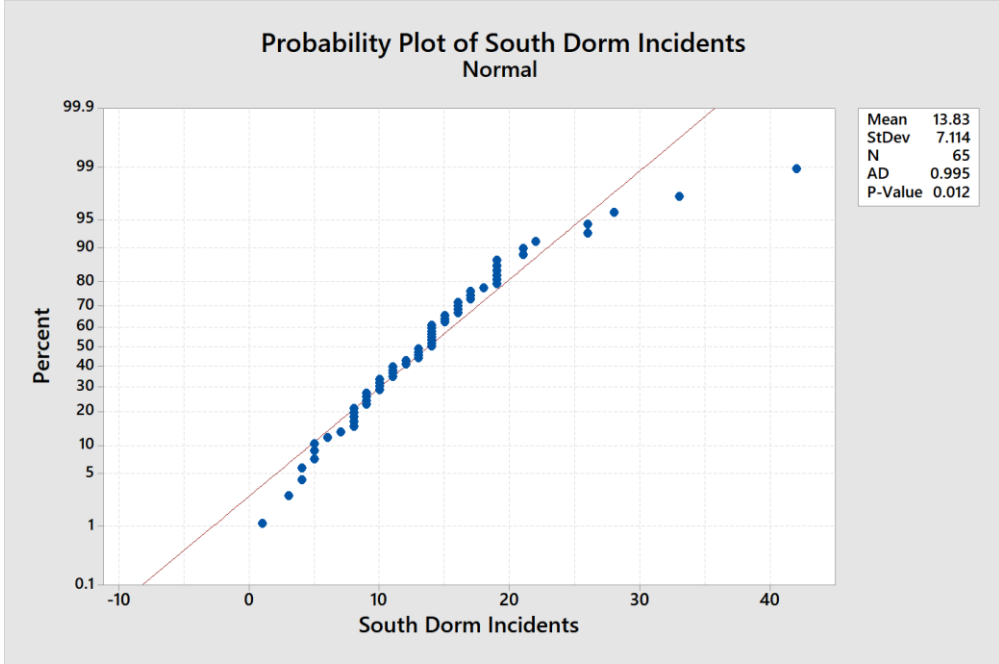


Heatmap of All Incidents in Student Housings only

Sources: Google Maps.

North Dorms vs. South Dorms?





Approximately Normally Distributed

Sources: Minitab.

North Dorms vs. South Dorms?

Hypothesis Test

$H_0: \mu_1 = \mu_2$ *vs.* $H_1: \mu_1 \neq \mu_2$

(μ_1 : Average number of incidents in North Dorms per month,
μ_2 : Average number of incidents in South Dorms per month)

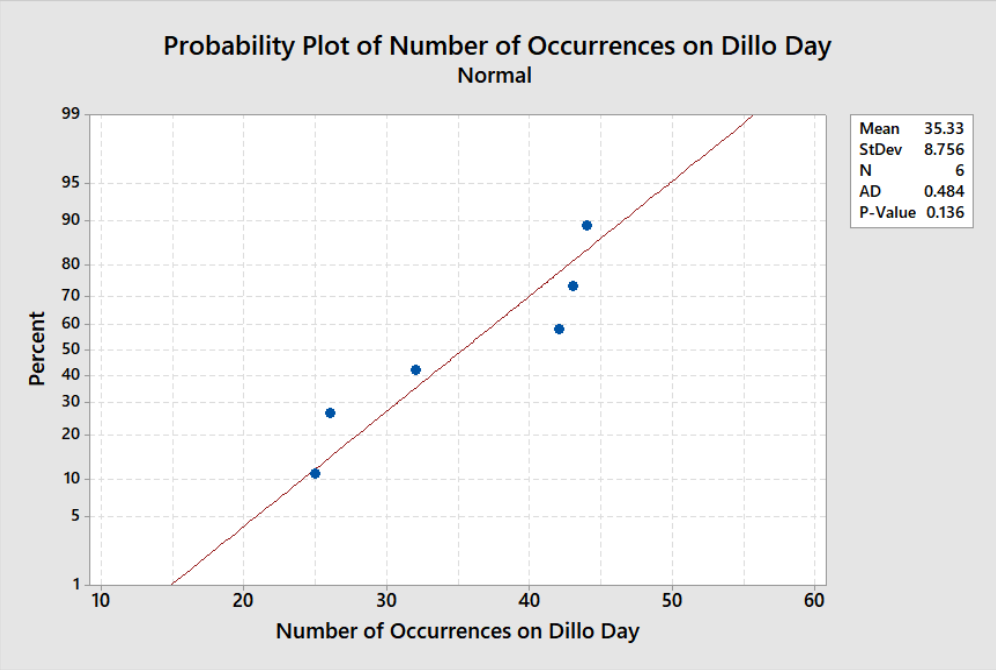
Two-Sample T-Test and CI: N, S

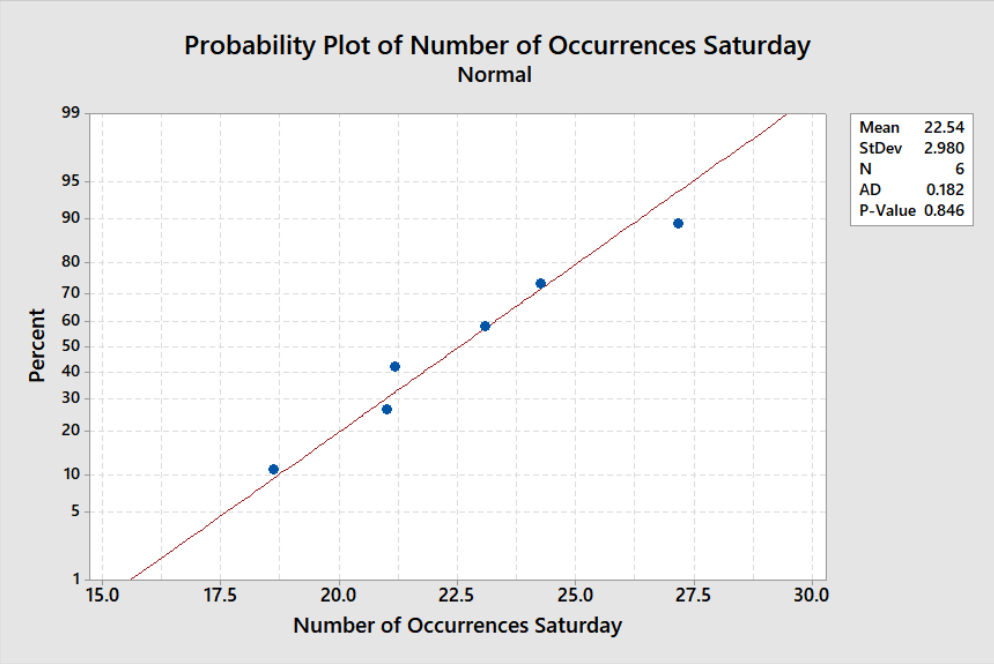
```
Two-sample T for N vs S

    N   Mean  StDev  SE Mean
N   65  10.57   6.52     0.81
S   65  13.83   7.11     0.88


Difference = μ (N) - μ (S)
Estimate for difference:  -3.26
95% CI for difference:  (-5.63, -0.89)
T-Test of difference = 0 (vs ≠): T-Value = -2.72  P-Value = 0.007  DF = 127
```

Reject H_0 ; Statistically significant difference

Sources: Minitab.

How crazy are Dillo Days?





Approximately Normally Distributed

Sources: Minitab.

How crazy are Dillo Days?

Hypothesis Test

$H_0: \mu_1 = \mu_2$ *vs.* $H_1: \mu_1 \neq \mu_2$

(μ_1 : Average number of incidents on each Dillo Day,
μ_2 : Average number of incidents on regular Saturdays)

Two-Sample T-Test and CI: Number of Occurances on Dillo D, Number of Occurances Saturday

```
Two-sample T for Number of Occurances on Dillo D vs Number of Occurances Saturday

                            N   Mean  StDev  SE Mean
Number of Occurances on     6  35.33   8.76      3.6
Number of Occurances Sat    6  22.54   2.98      1.2


Difference = μ (Number of Occurances on Dillo D) - μ (Number of Occurances Saturday)
Estimate for difference:  12.79
95% CI for difference:  (3.55, 22.03)
T-Test of difference = 0 (vs ≠): T-Value = 3.39  P-Value = 0.015  DF = 6
```

Reject H_0 ; Statistically significant difference

Sources: Minitab.

How is each day of the week different?

Hypothesis Test

H_0: $\mu_1 = \mu_2 = \ldots = \mu_7$ *vs.*

H_1: *at least one* μ_i *is different*

(where each μ_i represents the average number of incidents on a day of the week, Monday through Sunday)

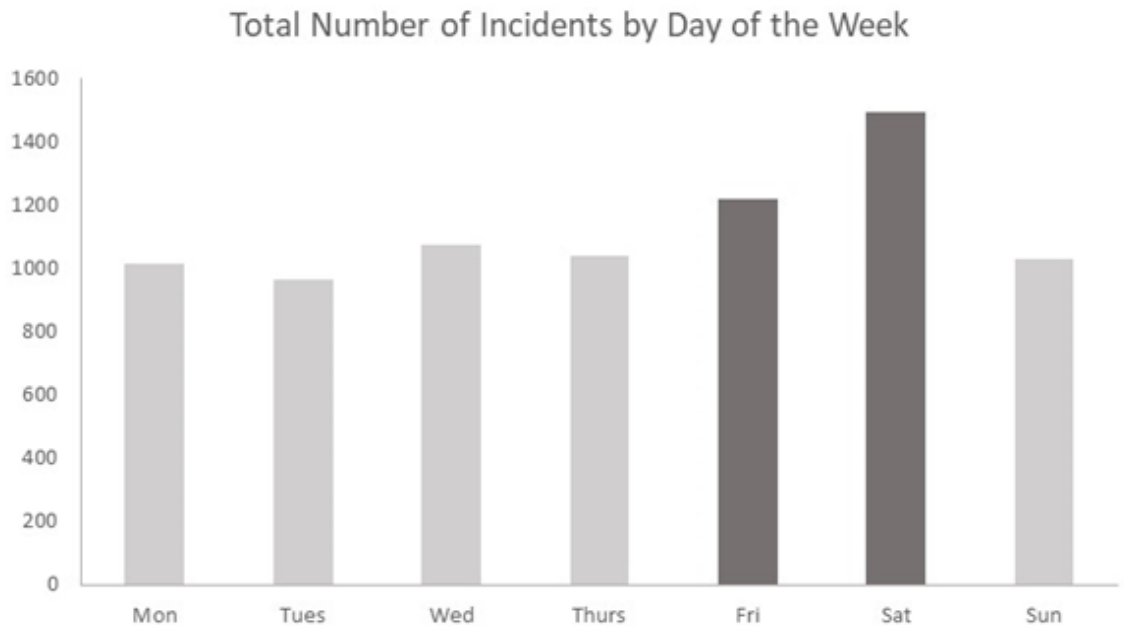
Reject H_0 ; Statistically significant difference

```
Call:
lm(formula = Count ~ ., data = DAY)

Residuals:
    Min      1Q  Median      3Q     Max
-21.969  -5.254  -0.754   3.354  45.031

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.8308     1.0862  14.575  < 2e-16 ***
Mon          -0.1846     1.5361  -0.120   0.9044
Tues         -1.0154     1.5361  -0.661   0.5089
Wed           0.7077     1.5361   0.461   0.6452
Thurs         0.1385     1.5361   0.090   0.9282
Fri           2.9231     1.5361   1.903   0.0577 .
Sat           7.1385     1.5361   4.647 4.43e-06 ***
Sun               NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.757 on 448 degrees of freedom
Multiple R-squared:  0.08266,   Adjusted R-squared:  0.07037
F-statistic: 6.728 on 6 and 448 DF,  p-value: 7.752e-07

Analysis of Variance Table

Response: Count
           Df Sum Sq Mean Sq F value   Pr(>F)
Mon         1    187  187.26  2.4420  0.11883
Tues        1    554  553.60  7.2194  0.00748 **
Wed         1    176  176.49  2.3016  0.12995
Thurs       1    504  504.01  6.5727  0.01068 *
Fri         1     18   18.09  0.2359  0.62739
Sat         1   1656 1656.12 21.5971 4.43e-06 ***
Residuals 448  34354   76.68
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sources: R.

How is each day of the week different?

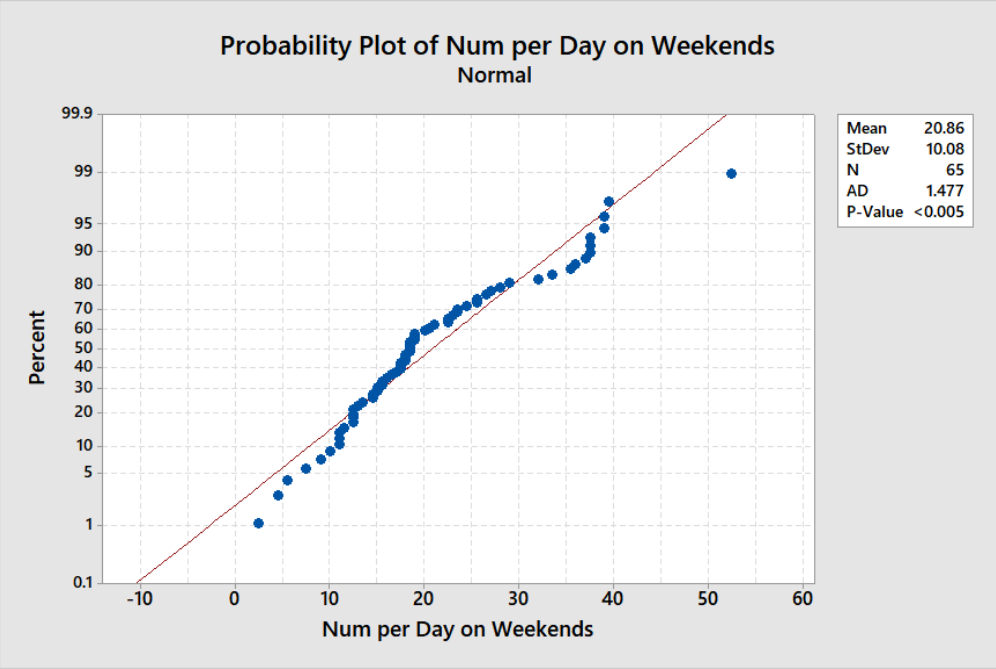
Grouping Weekend vs. Weekday

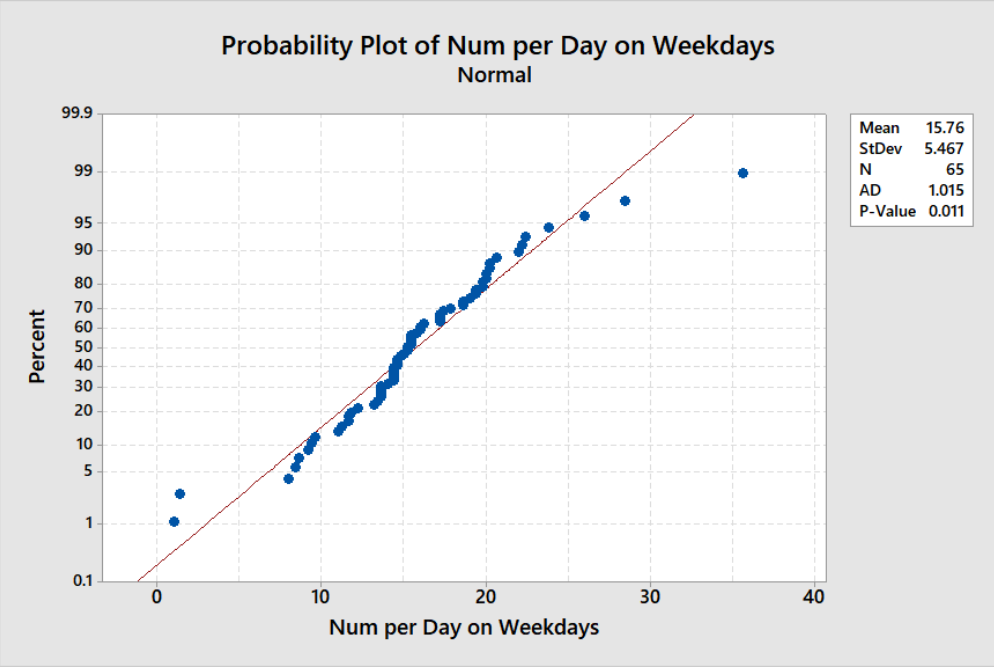



```
Call:
lm(formula = Count ~ ., data = DAY)

Residuals:
    Min      1Q  Median      3Q     Max
-21.969  -5.254  -0.754   3.354  45.031

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.8308     1.0862  14.575  < 2e-16 ***
Mon          -0.1846     1.5361  -0.120   0.9044
Tues         -1.0154     1.5361  -0.661   0.5089
Wed           0.7077     1.5361   0.461   0.6452
Thurs         0.1385     1.5361   0.090   0.9282
Fri           2.9231     1.5361   1.903   0.0577 .
Sat           7.1385     1.5361   4.647 4.43e-06 ***
Sun               NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.757 on 448 degrees of freedom
Multiple R-squared:  0.08266,   Adjusted R-squared:  0.07037
F-statistic: 6.728 on 6 and 448 DF,  p-value: 7.752e-07
```

Sources: R.

How is each day of the week different?





Approximately Normally Distributed

Sources: Minitab.

How is each day of the week different?

Hypothesis Test

$H_0: \mu_1 = \mu_2$ *vs.* $H_1: \mu_1 \neq \mu_2$

(μ_1 : Average number of incidents on weekends per day,
μ_2 : Average number of incidents on weekdays per day)

Two-Sample T-Test and CI: Num per Day on Weekends, Num per Day on Weekdays

```
Two-sample T for Num per Day on Weekends vs Num per Day on Weekdays

                           N   Mean  StDev  SE Mean
Num per Day on Weekends   65   20.9   10.1      1.3
Num per Day on Weekdays   65  15.76   5.47     0.68


Difference = μ (Num per Day on Weekends) - μ (Num per Day on Weekdays)
Estimate for difference:  5.10
95% CI for difference:  (2.28, 7.92)
T-Test of difference = 0 (vs ≠): T-Value = 3.59 P-Value = 0.001  DF = 98
```

Reject H_0 ; Statistically significant difference

Sources: Minitab.

Multiple Regression: Incident Rate

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	155.593	22.2275	15.23	0.000
North	1	3.705	3.7051	2.54	0.112
Weekday	1	12.883	12.8832	8.83	0.003
Away	1	37.231	37.2314	25.52	0.000
North*Weekday	1	0.074	0.0741	0.05	0.822
North*Away	1	0.008	0.0080	0.01	0.941
Weekday*Away	1	1.162	1.1616	0.80	0.373
North*Weekday*Away	1	0.018	0.0180	0.01	0.912
Error	252	367.693	1.4591		
Total	259	523.286			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.20793	29.73%	27.78%	25.66%

Coefficients

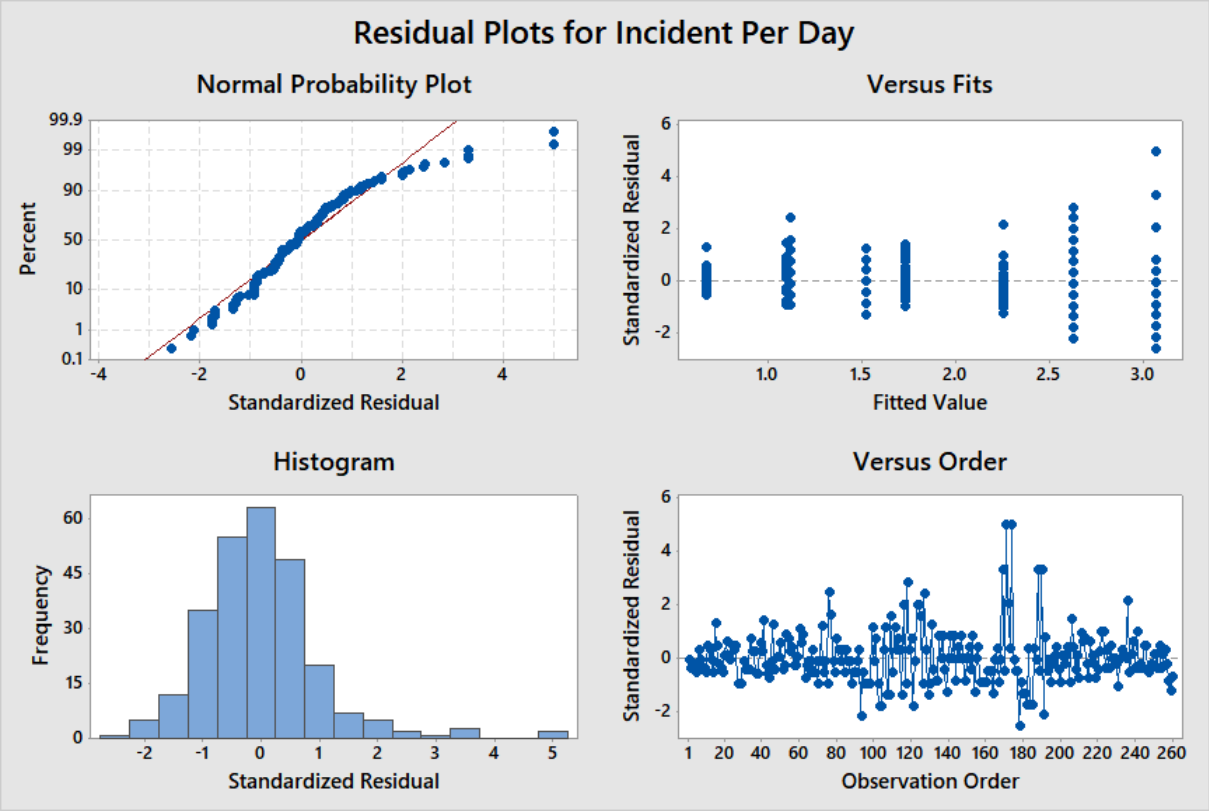
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.064	0.193	15.84	0.000	
North	-0.436	0.274	-1.59	0.112	3.33
Weekday	-0.813	0.274	-2.97	0.003	3.33
Away	-1.545	0.306	-5.05	0.000	4.00
North*Weekday	-0.087	0.387	-0.23	0.822	5.00
North*Away	0.032	0.433	0.07	0.941	5.33
Weekday*Away	0.386	0.433	0.89	0.373	5.33
North*Weekday*Away	0.068	0.612	0.11	0.912	6.00

Sources: Minitab.

Incident Per Day = 3.064 – 0.436*North – 0.813*Weekday – 1.545*Away – 0.087*North*Weekday + 0.032*North*Away + 0.386*Weekday*Away + 0.068*North*Weekday*Away

Multiple Regression: Incident Rate

Incident Per Day = 3.064 – 0.436*North – 0.813*Weekday – 1.545*Away – 0.087*North*Weekday + 0.032*North*Away + 0.386*Weekday*Away + 0.068*North*Weekday*Away



Sources: Minitab.

Multiple Regression: Incident Rate

Incident Per Day = 3.064 – 0.436*North – 0.813*Weekday – 1.545*Away – 0.087*North*Weekday + 0.032*North*Away + 0.386*Weekday*Away + 0.068*North*Weekday*Away

Data Type	95% Confidence Intervals		95% Prediction Intervals	
	Lower	Upper	Lower	Upper
(North, Weekend, Students present)	2.25	3.01	0.22	5.04
(North, Weekend, Students away)	0.65	1.58	-1.31	3.54
(North, Weekday, Students present)	1.35	2.11	-0.68	4.14
(North, Weekday, Students away)	0.20	1.14	-1.76	3.09
(South, Weekend, Students present)	2.68	3.45	0.65	5.47
(South, Weekend, Students away)	1.06	1.99	-0.91	3.94
(South, Weekday, Students away)	0.63	1.56	-1.33	3.52
(South, Weekday, Students present)	1.87	2.63	-0.16	4.66

Sources: Minitab.

Conclusion

- Overall summary of our findings
  - Statistically significant differences between incidents rate on dorms in North vs South campus; weekend vs weekdays; Dillo vs regular Saturdays; months students are on campus vs months students are away on break
  - Multiple regression on incident rate on campus
- Limitations
- Possible improvements for the future

Questions & Answers

How do different months compare?

Hypothesis Test

H_0: $\mu_1 = \mu_2 = \ldots = \mu_{12}$ *vs.*

H_1: *at least one* μ_i *is different*

(where each μ_i represents the average number of incidents on different months, January through December)

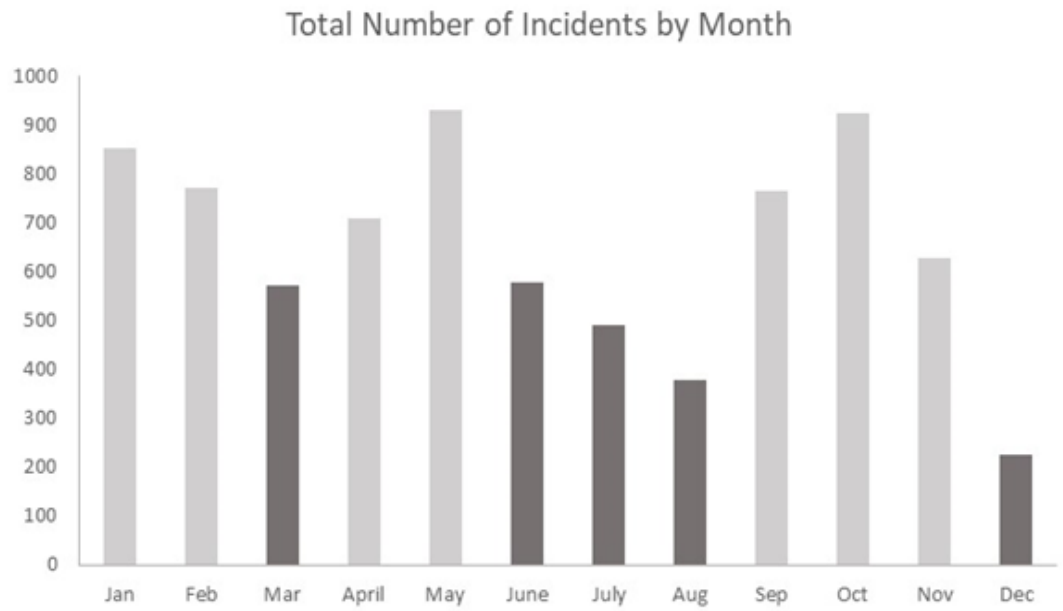
Reject H_0 ; Statistically significant difference

```
Call:
lm(formula = Count ~ Jan + Feb + Mar + Apr + May + Jun + Jul +
    Aug + Sep + Oct + Nov, data = MONTH)

Residuals:
   Min     1Q Median     3Q    Max
-57.33 -12.40  -2.50   9.80 110.83

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    45.20      12.38   3.651 0.000598 ***
Jan            96.97      16.76   5.785 3.98e-07 ***
Feb            83.63      16.76   4.989 6.90e-06 ***
Mar            50.30      16.76   3.001 0.004100 **
Apr            72.80      16.76   4.343 6.37e-05 ***
May           110.13      16.76   6.570 2.22e-08 ***
Jun            70.40      17.51   4.021 0.000185 ***
Jul            53.20      17.51   3.039 0.003686 **
Aug            30.60      17.51   1.748 0.086298 .
Sep           108.00      17.51   6.169 9.75e-08 ***
Oct           139.80      17.51   7.985 1.18e-10 ***
Nov            80.60      17.51   4.604 2.63e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.68 on 53 degrees of freedom
Multiple R-squared:  0.6674,    Adjusted R-squared:  0.5983
F-statistic: 9.667 on 11 and 53 DF,  p-value: 2.847e-09
```

Sources: R.

How do different months compare?

Grouping Months Students are present vs. away




```
Call:
lm(formula = Count ~ Jan + Feb + Mar + Apr + May + Jun + Jul +
    Aug + Sep + Oct + Nov, data = MONTH)

Residuals:
   Min     1Q Median     3Q    Max
-57.33 -12.40  -2.50   9.80 110.83

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    45.20      12.38   3.651 0.000598 ***
Jan            96.97      16.76   5.785 3.98e-07 ***
Feb            83.63      16.76   4.989 6.90e-06 ***
Mar            50.30      16.76   3.001 0.004100 **
Apr            72.80      16.76   4.343 6.37e-05 ***
May           110.13      16.76   6.570 2.22e-08 ***
Jun            70.40      17.51   4.021 0.000185 ***
Jul            53.20      17.51   3.039 0.003686 **
Aug            30.60      17.51   1.748 0.086298 .
Sep           108.00      17.51   6.169 9.75e-08 ***
Oct           139.80      17.51   7.985 1.18e-10 ***
Nov            80.60      17.51   4.604 2.63e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 27.68 on 53 degrees of freedom
Multiple R-squared:  0.6674,    Adjusted R-squared:  0.5983
F-statistic: 9.667 on 11 and 53 DF,  p-value: 2.847e-09
```

Sources: R.

How do different months compare?

Hypothesis Test

$H_0: \mu_1 = \mu_2$ *vs.* $H_1: \mu_1 \neq \mu_2$

(μ_1 : Average number of incidents in months students are away,
μ_2 : Average number of incidents in months students are here)

Two-Sample T-Test and CI: Num Per Month on Break, Num Per Month During Classes

```
Two-sample T for Num Per Month on Break vs Num Per Month During Classes

                          N   Mean  StDev  SE Mean
Num Per Month on Break    5  85.84   3.82      1.7
Num Per Month During Cla  5  147.3   14.3      6.4

Difference = μ (Num Per Month on Break) - μ (Num Per Month During Classes)
Estimate for difference:  -61.50
95% CI for difference:  (-79.92, -43.09)
T-Test of difference = 0 (vs ≠): T-Value = -9.27 P-Value = 0.001 DF = 4
```

Reject H_0 ; Statistically significant difference

Sources: Minitab.