Analysis of fixed effects on Garage Parking Space at San José State Univeristy

Department of Mathematics and Statistics

San José State University

By

Kim, Jung-a

Sun, Liyuan

Yin, Mengqi

May, 2019

1. Executive Summary

The purpose of our research is to find the right time to arrive at school to find parking space in the garages at San Jose State university. Every semester, 20,000 students which are about 41% of total students drive to school and there are only about 5,000 parking spots including Employees section on main campus and 850 spots near Spartan Stadium. Thus, students have to compete for spots by arriving earlier than their first class schedule or find alternatives. Sameer Saran, SJSU MS Computer Engineering 18' and Hooman Bolandi launched an app called 'Parkstash' to tackle this problem by connecting students to paid parking in the neighborhood of the university. Parkstash also provides the proportion of parked vehicles in percentage.

We collected ratio of parked vehicles 10 and 30 minutes prior to beginning of each class time in each garage on campus and Park&Ride near the SJSU stadium during the first week of the final month of Spring semester. The ratio included all the parking spots in each garage including General, Employees, Disabled, Housing, R-permits, and 20-minute zones. Commuting students can park in only General section. The proportion of General parking space was about 78% in North and South garages. West garage had 99% of General parking space. Park&Ride did not have any parking restrictions. Thus, we took this into account in our conclusions since Employee parking space are usually emptier than General parking space.

The data were analyzed to answer our questions such as whether arriving early ahead of class schedule helps students find parking space with high probability and whether they should find alternatives if they arrive close to their class schedule. It can be used to predict the demand in private parking space near SJSU and carpooling service according to the significant factors. Furthermore, we can use the results as the evidence to claim the necessity of restriction on the sales volume of parking permits or building a new parking structure on campus.

2. Factors

We used hypothesis testing to find the significant factors. The statistical null hypotheses were "Class times/Arriving times/Day does not have significant effect on ratio of parked vehicles in a garage at SJSU." and "There are no interaction effects among these factors on the ratio." There were three fixed factors: Class Time, Arrival Time, and Class Days. The observational units were the four garages: Park&Ride, South, North, and West garages. The observations were the proportion of parked vehicles in the garage. The ratio were repeatedly measured at 30 minutes and 10 minutes prior to each of 5 class beginning times which are 9:00AM, 10:30, 12:00, 1:30, and 3:00PM on weekdays(Monday through Thursday). Each pair of weekdays(Monday/Wednesday and Tuesday/Thursday) were decided as the only levels of Class Days since class schedules are the same for two days. Our full ANOVA model was three-way fixed effects model with a blocking variable.

$$y_{ijklm} = \delta_i + \alpha_j + \beta_k + \gamma_l + \alpha\beta_{jk} + \beta\gamma_{kl} + \alpha\gamma_{il} + \alpha\beta\gamma_{jkl} + \varepsilon_{ijklm} \text{ where m = 1, 2}$$

The garage was selected as the blocking variable(δ) where data were repeatedly collected. There were two observations per combination. The factor levels were encoded as follows.

Variable	Garage	Class Time	Arrival Time	Class Days
Levels coding	P = Park&Ride S = South N = North	1 = 9:00-10:15 2 = 10:30-11:45 3 = 12:00-1:15	30 = 30 minutes 10 = 10 minutes	MW = Monday/Wednesday TTh = Tuesday/Thursday

	W = West	4 = 1:30-2:45 5 = 3:00-4:15		
Number of levels	4	5	2	2

3. Model

Fitting the full model, we noticed a quadratic pattern in residual plots against Garage and Class Time. We used Box-Cox method to transform the response and found the maximum likelihood estimate of lambda at 1.785. We squared the response and re-examined the normal qq-plot and the residual plots against each factor. The residual plots and qq-plot satisfied the assumptions(Figure 1). There were three outliers that had very low parking ratio. These were observed at 8:30AM in Park&Ride. Two of them were on Tuesday/Thursday and the other one was on Monday/Wednesday. This was assumably because the garages on campus had plenty of parking spots around 50% at 8:30AM so students usually would have gone to campus parking lots unless they had Park&Ride-only permits.

After the transformation, we looked at p-values of each factor's F-test statistics through JMP output.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Garage	3	3	376658511	563.3006	<.0001*
Class Time	4	4	616744089	691.7652	<.0001*
Class Days	1	1	442892	1.9871	0.1609
Arrival Time	1	1	7207161	32.3354	<.0001*
Class Time*Class Days*Arrival Time	4	4	107156	0.1202	0.9751
Class Time*Class Days	4	4	293575	0.3293	0.8579
Class Days*Arrival Time	1	1	348	0.0016	0.9685
Class Time*Arrival Time	4	4	13949237	15.6460	<.0001*

Table 1. F-tests of factors

From Table 1, Class Days was clearly not a significant factor. Thus, interaction terms involving Class Days were not significant with huge p-values. Removing the four terms that involved Class Days, the model made two-way fixed effects model $y_{ijkl} = \delta_i + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijkl}$. As expected, Class Time and Arrival Time were critical in determining the parking ratio.

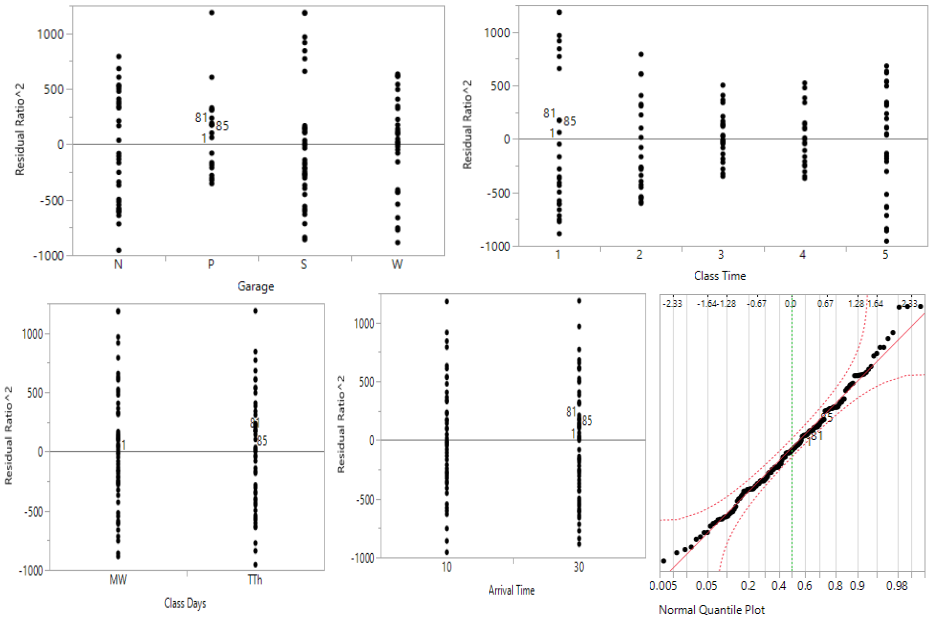
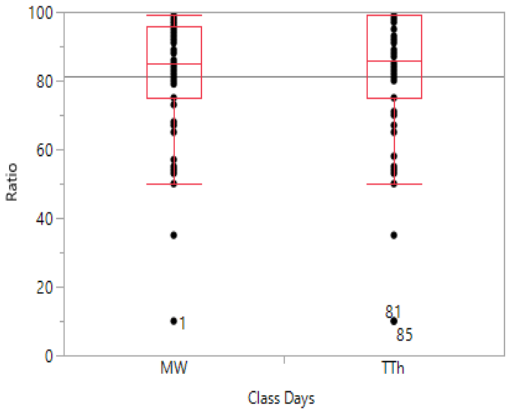


Figure 1. Residual plots and qq-plot of squared model

4. Class Days



Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	0.1564	1	158	0.6931
Brown-Forsythe	0.0974	1	158	0.7554
Levene	0.1247	1	158	0.7245
Bartlett	0.3803	1	.	0.5374
F Test 2-sided	1.1494	79	79	0.5374

Pooled t Test

TTh-MW

Assuming equal variances

Difference	0.3500	t Ratio	0.117972
Std Err Dif	2.9668	DF	158
Upper CL Dif	6.2097	Prob > |t|	0.9062
Lower CL Dif	-5.5097	Prob > t	0.4531
Confidence	0.95	Prob < t	0.5469

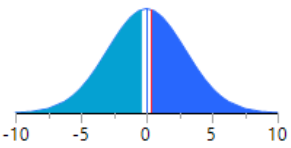
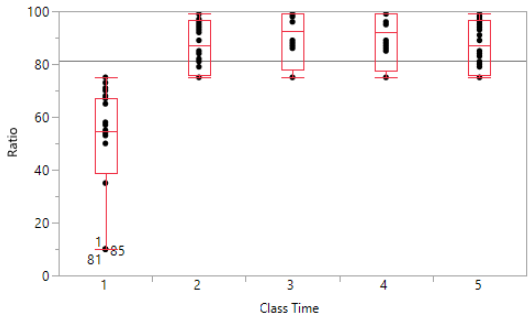


Figure 2. Class Days

The two-sided F-test statistic supported that the variances were equal with p-value = 0.5374. Thus, we could use t-test with pooled variance to test whether the means were equal between the pairs of Class Days. The two-tailed t-test statistic showed extremely large p-value(0.9062), so we retained the null hypothesis that the means were equal.

5. Class Time



HSD Threshold Matrix

Abs(Dif)-HSD

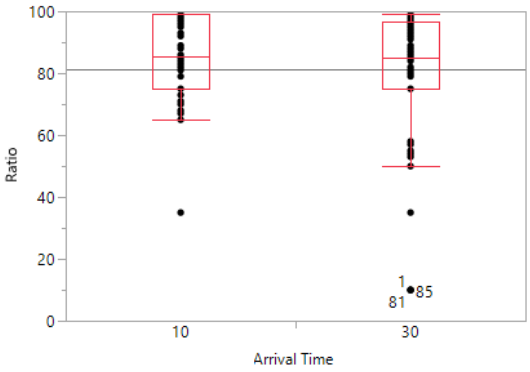
	3	4	2	5	1
3	-8.177	-8.021	-5.302	-5.115	29.573
4	-8.021	-8.177	-5.459	-5.271	29.416
2	-5.302	-5.459	-8.177	-7.990	26.698
5	-5.115	-5.271	-7.990	-8.177	26.510
1	29.573	29.416	26.698	26.510	-8.177

Positive values show pairs of means that are significantly different.

Figure 3. Boxplots of Class Time and Tukey-Kramer HSD

There was a noticeable difference in variance between the first level and the others from the box plot. All the data points including the outliers were below the overall average at the first level. The other levels seemed to have similar average and variance, but there was a slight quadratic pattern that slowly decreased after third level. The mean of first level was 52.28 while the others were above 86. Using Tukey-Kramer method, we could see that the difference between the means of the first level and any other level was much greater than HSD with overall desired significance level at 0.05. Thus, we concluded that the parking space ratio was significantly lower for 9:00AM class than any other class time.

6. Arrival Time and Class Time

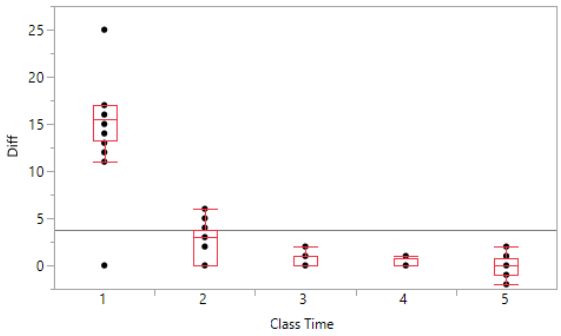



```
        One Sample t-test

data:  difference
t = 4.9944, df = 79, p-value = 3.458e-06
alternative hypothesis: true mean is not equal to
0
95 percent confidence interval:
 2.255484 5.244516
sample estimates:
mean of x
     3.75
```

Figure 4. Boxplot and output of R function `t.test`(One sample t-test with differences) for Arrival Time

Considering that there was a significant interaction between Arrival Time and Class Time, it was evident that parking ratio at Arrival Time levels for the same Class Time level would be similar. Thus, a paired-t test was conducted as it removed the effect of Class Time.



Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
1	16	15.6875	6.1070315	1.5267579	12.433293	18.941707
2	16	2.4375	1.9989581	0.4997395	1.3723304	3.5026696
3	16	0.5625	0.7274384	0.1818596	0.1748754	0.9501246
4	16	0.25	0.4472136	0.1118034	0.0116967	0.4883033
5	16	-0.1875	1.1086779	0.2771695	-0.778273	0.4032727

Figure 5. Boxplot and ANOVA of difference between two Arrival times according to Class Time levels

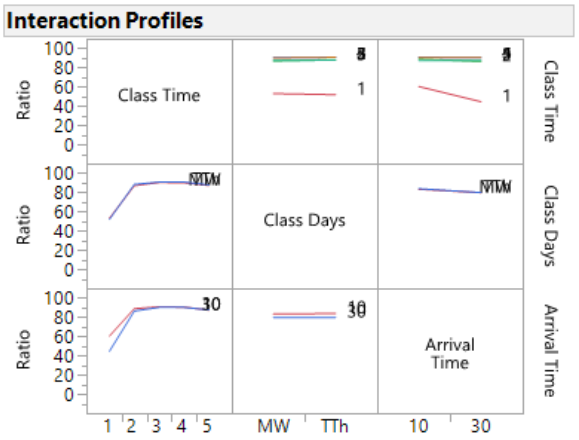


Figure 6. Interaction Plot among Class Time, Class Days, and Arrival Time

The paired t-test with differences between arrival times showed close-to-zero p-value in the R output(Figure 4), which meant there was a significant difference in parking space between the two arrival times on average.

The box plot(Figure 5) showed that the difference of ratio in Arrival Times was very high(10-17%) at the first level of Class Time. There were three outliers with 25% and one outlier with 0% at the first level. Since there were 16 observations per Class Time, the four outliers accounted for ¼ of the group size which was considerably a large proportion. Therefore, the standard deviation at the first level was more than three times higher than the other levels. After the first level, the difference in parking space sharply decreased to below 5% with much smaller variance. This could explain the interaction plot between Class Time and Arrival Time(Figure 6) where 30-minutes showed lower ratio than 10-minutes at the first level. There seemed to be no interaction in other combinations in the plot.

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
110	16	60.125	15.230999	3.8077498	52.008973	68.241027
130	16	44.4375	17.828699	4.4571747	34.937257	53.937743
210	16	88.375	10.005832	2.5014579	83.043269	93.706731
230	16	85.9375	8.8804561	2.220114	81.205439	90.669561
310	16	90.3125	10.130605	2.5326514	84.914281	95.710719
330	16	89.75	10.003333	2.5008332	84.4196	95.0804
410	16	90	10.092902	2.5232254	84.621872	95.378128
430	16	89.75	10.109402	2.5273504	84.36308	95.13692
510	16	86.875	9.5699181	2.3924795	81.775551	91.974449
530	16	87.0625	9.454937	2.3637342	82.02432	92.10068

Table 2. One way ANOVA of Class Time*Arrival Time

Table 2 shows the mean parking ratio in 10 combinations of each level of Class Time and Arrival Time(The first digit indicates the level of Class Time and the latter two digits indicate Arrival Time). Since 110 and 130 were noticeably lower than the other combinations, we conducted a pooled t-test to compare the means between the first level of Class Time and the others(we

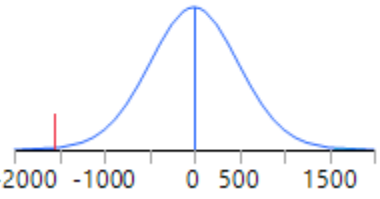
assumed equal variance since the two-sided F-test statistic p-value was 0.7484 and we used squared parking ratio to satisfy normality assumption with data.) The t-test p-value was less than 0.0001, so we concluded that there was a significant difference between the first level of Class Time and the others. Then, we conducted a pooled t-test to compare the two means of 110 and 130 to see if there was difference between the arrival times at the first level of Class Time(with the same assumptions as before).

Pooled t Test

130-110

Assuming equal variances

Difference	-1559.8	t Ratio	-3.14075
Std Err Dif	496.6	DF	30
Upper CL Dif	-545.5	Prob > |t|	0.0038*
Lower CL Dif	-2574.1	Prob > t	0.9981
Confidence	0.95	Prob < t	0.0019*



From figure 7, we could see that there was difference between 130 and 110, which meant parking space at 8:30AM was significantly lower than 8:50AM.

Figure 7. Pooled t-test between 130 and 110

7. Garage

Since garage was our observational unit/blocking variable in this model, we did not conduct F-tests for individual terms to test whether to include them in the model. However,

there seemed to be a significant difference among garages so we used Tukey-Kramer method to compare all of the pairs. At each garage, 40 ratios were observed.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
W	P	26.15000	3.520537	17.0073	35.29266	<.0001*
N	P	23.92500	3.520537	14.7823	33.06766	<.0001*
S	P	14.47500	3.520537	5.3323	23.61766	0.0004*
W	S	11.67500	3.520537	2.5323	20.81766	0.0062*
N	S	9.45000	3.520537	0.3073	18.59266	0.0398*
W	N	2.22500	3.520537	-6.9177	11.36766	0.9216

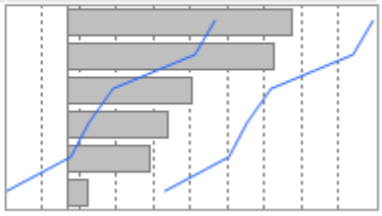


Figure 8. Tukey-Kramer HSD for garages

Using $q_{0.05;4,156} \approx 2.59695$, the test statistic for difference between West and North garages was $\frac{2.22500}{3.520537} < 2.59695$, thus we retained the null hypothesis that there was no difference in parking space between West and North garage. All the other pairs had a significant difference.

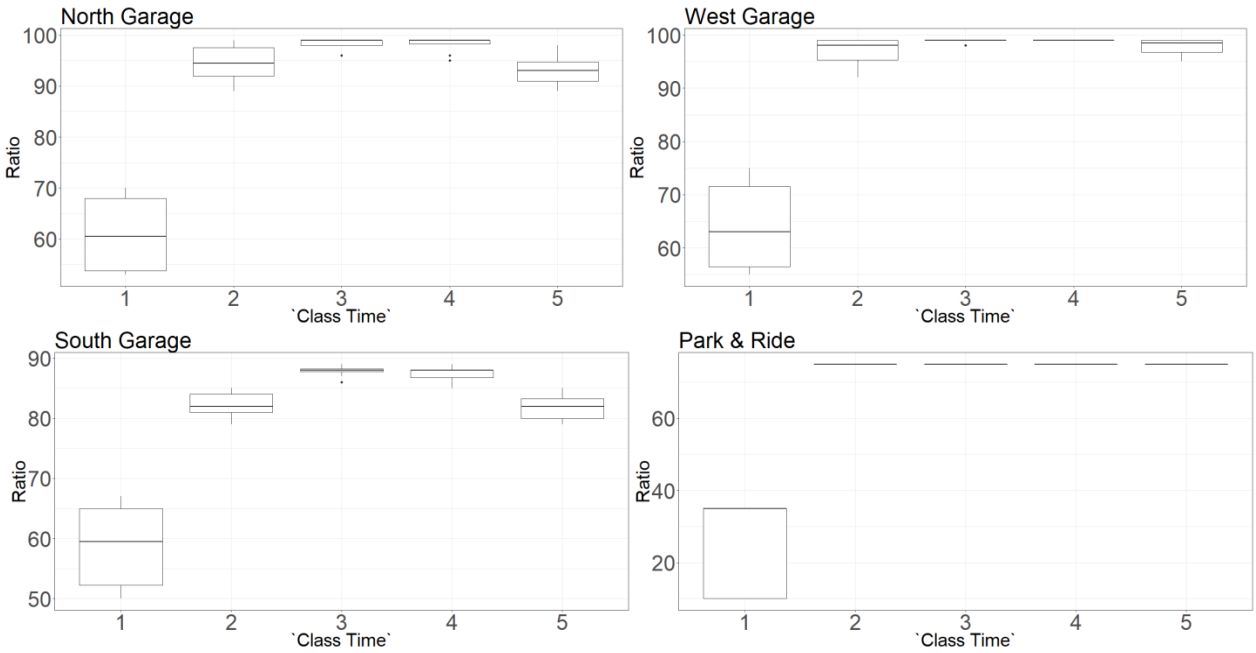


Figure 9. Boxplots of garages against Class Time

Since there was a significant difference among garages, we were curious whether Class Time affects ratio at each garage differently. From the box plot against Class Time(Figure 9), we could easily see that the distributions of North and West garages were almost the same maintaining above 90% except for the first level of Class Time. West garage was slightly more dramatic as the ratio was 99% at 23 out of total 40 runs and almost full until the fifth level while North garage dropped slightly to around 95% after the fourth level. South garage was less busier than North and West garages throughout all five levels as the ratio was below 90% at all times. In Park&Ride garage, the parking ratio were 35% or 10% before the first level and maintained exactly 75% at the rest of the levels.

8. Conclusion

Since this was an observational study, we could not conclude that Class Time, Arrival Time, and Garage affect the parking space ratio. However, in the ordinary days without particular events such as concerts or graduation ceremonies that invite people other than SJSU students and faculties, it would be hard to imagine a hidden variable that significantly affect parking space since students usually arrive at school according to their class schedules. Thus, although we cannot provide the evidence that the factors affect the parking space in this study, we propose that their relationship is rather causational than correlational with assumably little chance that such hidden variables existed during the period of the experiment.

Class Time and Arrival Time were the main factors that affect parking ratio whereas there was no significant difference between Monday/Wednesday and Tuesday/Thursday schedules. For Class Time, there was a significant difference in parking ratio between the first level and the other levels based on the Tukey-Kramer HSD result. In the box plot(Figure 3), we could see the variance at the first level was large whereas the variance was small for all the other levels with similar distributions. This was also related to the interaction term between Class Time and Arrival Time since parking ratio at 8:30AM and 8:50AM were significantly different. On the other hand, the difference was minimal for the other Class Time levels.

Although garage was our observational units, we could deduce many things from the boxplot of garages against Class Time with the knowledge of designated parking areas. Before the first level of Class Time, parking space ratio was below 75% at all four garages and even below 35% for Park&Ride(Figure 9). North and West garages had similar pattern although West was mostly full except for 9:00AM. Considering that 99% of West garage is General parking space, we assumed that the slight decrease at 1:00-1:20PM(fourth level) in North and South garages mainly attribute to employees leaving the campus since North and South garages have 19% of Employees space since it would be hard to justify that students who parked at North or South garage leave earlier than those who parked at West garage. Plus, the consistent 75% ratio of parking space at Park&Ride where mostly students park supports that most students do not leave campus until 3:00PM.

The students who have class at 9:00AM would find the parking spot in any of the four garages even though the space fill up quickly between 8:30 and 8:50 because there are fewer classes starting at 9:00AM than other schedules. But after 10:00AM, North and West garages are almost full around 95% and South garage is about 82% full. Thus, it would be risky enough for students who arrive at 10:00AM to find a parking spot on campus. Also, regarding the fact that South garage restricts the parking space on the rooftop to only employees and housing permits until noon, commuting students would not be able to park in most of the empty space that are left in South garage. From 11:30AM to 1:20PM(2-4th level), while North and West garages are almost completely full, South garage is about 12% empty during that time window, so after 12:00PM, students would be able to park on the rooftop. After 2:30PM, North garage is about 5% empty but it is not certain that it would be General space. South garage would be about 17% empty, but it has the same problem. West garage is almost full until 2:50PM. These results lead us to suggest to students to arrive early before 9:00AM or go to South garage in the afternoon and in case it fails, they should arrive at least 30 minutes before class so that they have enough time to go park in Park&Ride lot and to ride on the shuttle bus to campus.

Even though the parking space available for students is less than the quarter of students who claim to commute to school, it is not completely impossible to find parking space. However, it is burdening for students to find space in the busy parking lot in the limited amount of time before class. This problem alerted many commuting students to find alternatives, but no clear solution for all students has been suggested yet. Unless new parking space is created, we have to keep looking for the general pattern in parking space availability to optimize the arrival times for students to organize their plans throughout the semester.

References:

Saran, Sameer., *Parkstash*, Version 1.6.0, 2019. *Apple App Store*, https://www.findparkstash.com/.

Montgomery, D. (2017), *Design and Analysis of Experiments, 9th Edition*., Hoboken, NJ: John Wiley & Sons, Inc.

Sullivan, M., "Section 1.2: Observational Studies versus Designed Experiments," Elgin Community College. [Online]. Available: https://faculty.elgin.edu/dkernler/statistics/ch01/1-2.html

Castillo, A., "San Jose State alumni create new app to alleviate lack of parking for thousands at university," abc7NEWS, April 3rd, 2019. [Online]. Available: https://abc7news.com/technology/sj-state-alumni-create-new-app-to-alleviate-lack-of-parking-for-thousands-at-university/5231534/

Charles W. Davidson College of Engineering, San José State University, "ParkStash, "The AirBnB of Parking"," November 28th, 2018. [Online]. Available: https://engineering.sjsu.edu/news-and-events/news/parkstash-airbnb-parking

JMP®, Version <14.0.0>. SAS Institute Inc., Cary, NC, 1989-2019.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.