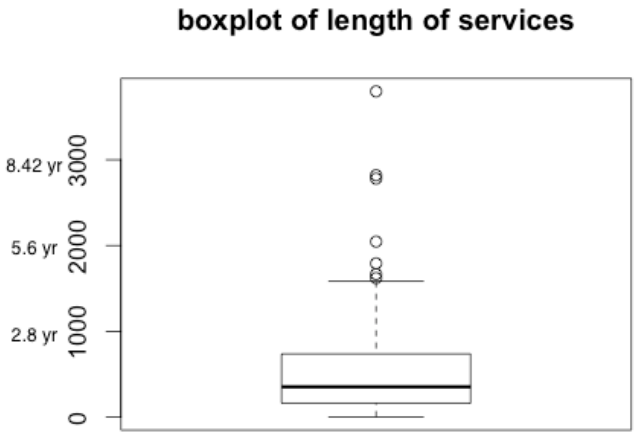
336 Project

Analysis of the effects of gender and age on elders' services

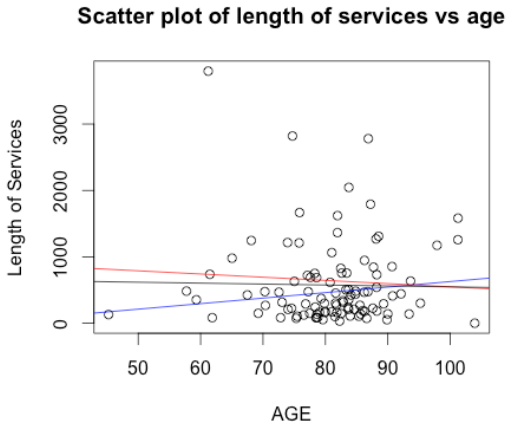
In this section we are looking at the impacts of age on the elder's services, such as length of service, days per week of care and number of caregivers separately. Whether gender makes a difference in the services is also a major study.

Length of service



From the boxplot of length of services, we can see that there are more than 5 outliners of elders who receive services from 3rd street for more than 4 years. Some of them even stay in 3rd Street Alliance for more than 10 years since the maximum is 3797.0 days. The longer the service, the better the elder's health is since the short length of service may imply that the elder passed away. In this situation we should focus more on the median instead of the mean since the outliers are too large and those outliers is the reason for the big gap between the median (353.5 days) and mean (574.1 days). Since the length of services is extreme positively skewed, a large amount of people die or transfer to a hospital due to their worse health within 1000 days (3 years).

Age VS Length of service



The overall regression doesn't show a relationship between age and length of services. The estimated regression line is $\hat{y} = 688.977 - 1.420 * x_i$ where $\hat{y}$ is the estimated length of service and xi is the age at intake. The slope is close to 1 so that the length of services does not change due to the age. Thus we have to look at genders.

The blue line represents a positive relationship between length of services and age for men in the sample. The estimated regression line is $\hat{y} = -203.802 + 8.319 * x_i$ where $\hat{y}$ is the estimated length of service and xi is the male's age at intake. The positive slope of 8.319 indicates a positive impact of age on length of service, which the older men stay in the Third Street Alliance longer. It means that as the male's age at intake increases by 1, his length of service will increase by 8.318 days.

But for women in the sample, there is a small negative relationship between the length of services and the age at intake. The estimated regression line is $\hat{y} = 1032.435 - 4.855x_i$ where $\hat{y}$ is the estimated length of service and xi is female's age at intake. The negative slope of -4.855 means that as the female's age at intake increases by 1, her length of service will shorten by 4.855 days.

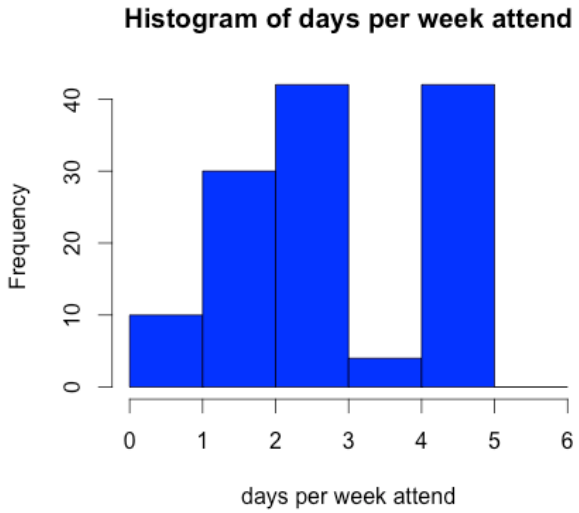
The difference between women and men may be caused by women's weaker immune system.

Test for hypothesis

Since we don't know the population variance of length of services, and the length of services is normally distributed by CLT, we use t test. Women and men are independent of each other, and then we view male and female as two populations. We did a two-sided t test for women and men. However based on the statistic we get, there is no enough evidence to conclude that the true mean of length of services for men is different from that of women. Even though the women's average length of services is much higher than that of men, the sample cannot support our hypothesis statistically due to some variability in this sample. The longest length of service of women is 3797 days (more than 10 years), but the longest length of service of men is 1621 days (approx 4.5 years). The large difference between the mean of women's length of services and that of men is because the number and magnitude of women's outliers are larger than those of men. The 95% CI suggest that the length of service difference between women and men will be between -395.7032 and 54.3064. So 95 percent of time, the true mean of difference between women and men's length of services will be in this interval.

```
t = -1.5053, df = 98.825, p-value = 0.1354
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -395.7032   54.3064
sample estimates:
mean of x mean of y
 468.6667  639.3651
```

Days per week of attend



It can be seen that weekly attendance has a bimodal distribution. Since it displays the frequencies of the number of days that elders come in for services, this may be indicative of two different levels of patient. The first mode may represent a group of more than 40 elders, which attend 3 days per week. The second representing those who attend 5 days per week. It also shows that elders rarely needed care only one day per week. Both the mean and the median are around 3 days per week. This may suggest that these individuals simply abstain from the program altogether.

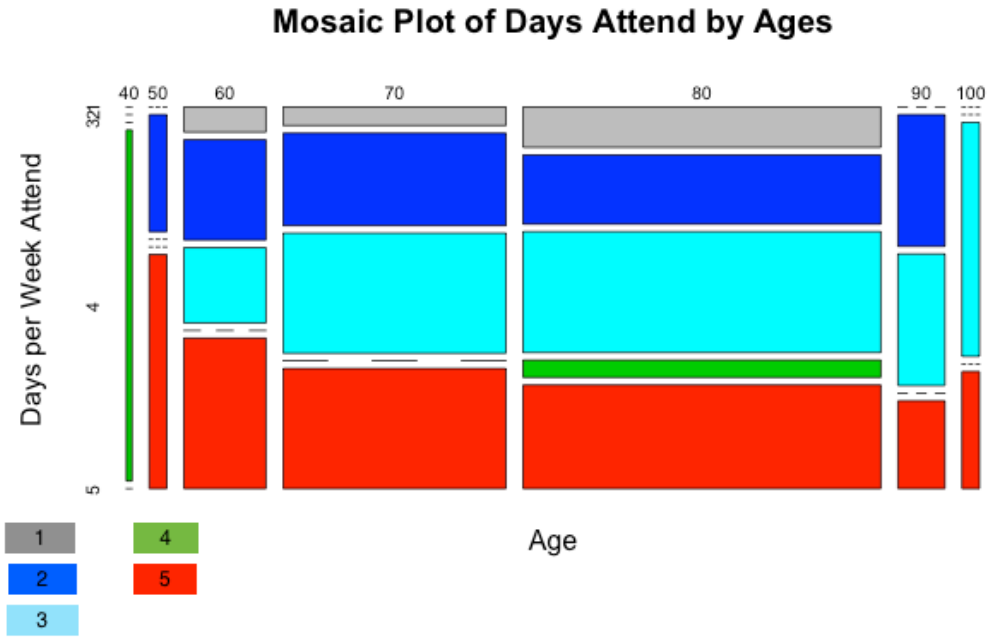


Figure 8 displays the relationship between age and days per week attend. Each color represents a specific attendance. The area of each bar reveals the number of elders in a specific age group with a specific attendance.

Out of 137 elders in Third Street Alliance, 128 people have recorded information of their attendances. The reason for the bars of 40s, 50s and 100s to be thin is that there are only 1 individual in 40s, 3 individuals in 50s and 3 individuals in 100s so that it is difficult for us to make a statistically conclusion of the relationship between age and days per week attend. What we can obtain from the plot is that only a small percentage of individuals attend once or four times a week since the areas of them in 60s, 70s, 80s and 90s are comparatively small. In addition, attending 3,4 or 5 times a week is common for almost each age group. The attendance of elders are sometimes decides by their family members since they usually send their parents to Third Street Alliance when they are busy with their work. Thus, if we intend to track those elders' health by their attendances, more information will be needed otherwise biases exist.

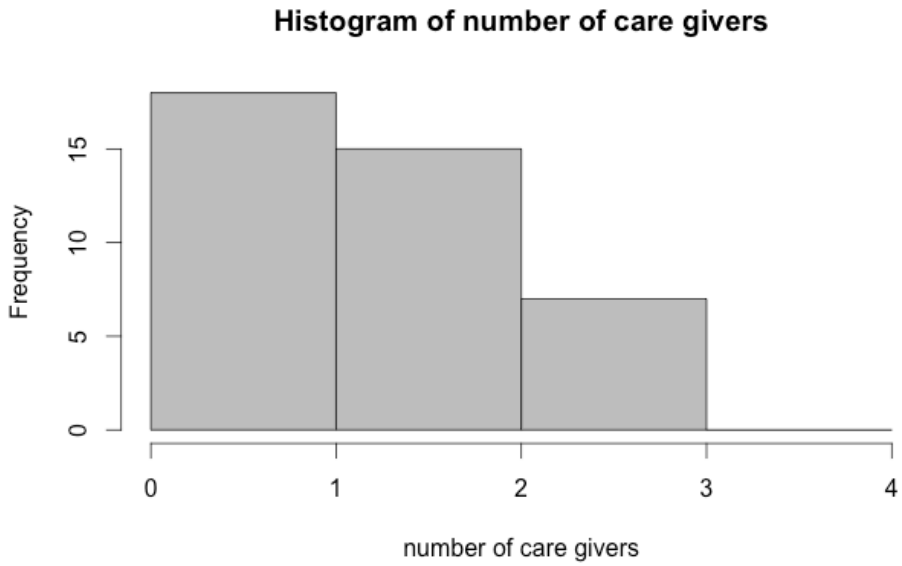
The spreads of different attendance are so different from each other even though they have very close medium age of elders. The spreads of once, twice, three times, five times per week are comparatively narrow while that of four times per week is wider. While elders attend three times a week are commonly aged from 70 to 98, there are four people who aged at 61, 65, 101 and 104 as outliers also attend 3 times per week. The youngest elder in record is aged at 45 and she attends 4 times a week. The distributions of once and four times are negatively skewed while the rest are normally distributed. Once and twice a week are the only attendances with outliers. Attendance of twice a week has an outlier of an elder aged at 60 while attendance of three times a week has outliers on both sides.

Test for hypothesis:

We view male and female as two populations (Dm, Df) for the days attend per week and we believe that there is a difference due to the gender. However, according to the t test we performed, the statistic is not significant enough to support our hypothesis, and the difference between the means is only 0.078. One possible reason could be that the elder's family members decide the attendance per week. Since children send their parents' to the 3rd Street Alliance due to their busy job duties or some other factors, the attendance is not heavily determined by the elders' health or wants. Thus, we cannot conclude that gender is a determinant for the days per week attends. Furthermore, the 95% CI suggests that the difference of days attend per week between women and men will be between -0.633 and 0.477. Thus, 95 percent of time the true difference in means of males' days per week attends and that of female will be in this interval.

```
t = -0.28039, df = 64.784, p-value = 0.7801
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6331001  0.4772268
sample estimates:
mean of x mean of y
 3.243902  3.321839
```

Number of caregivers



The distribution of the number of caregivers is unimodal and positive skewed. The histogram indicates that in this sample no one gets more than three caregivers. Out of 131 elders in the dataset, only 40 of them have recorded information their number of caregivers. Within those recorded information, everyone has at least 1 caregiver. The mode in the histogram may suggest that 1 or 2 caregivers will be enough for most elders since only 8 elders have 3 caregivers and no one has more than 3 caregivers.

Age VS Number of caregivers

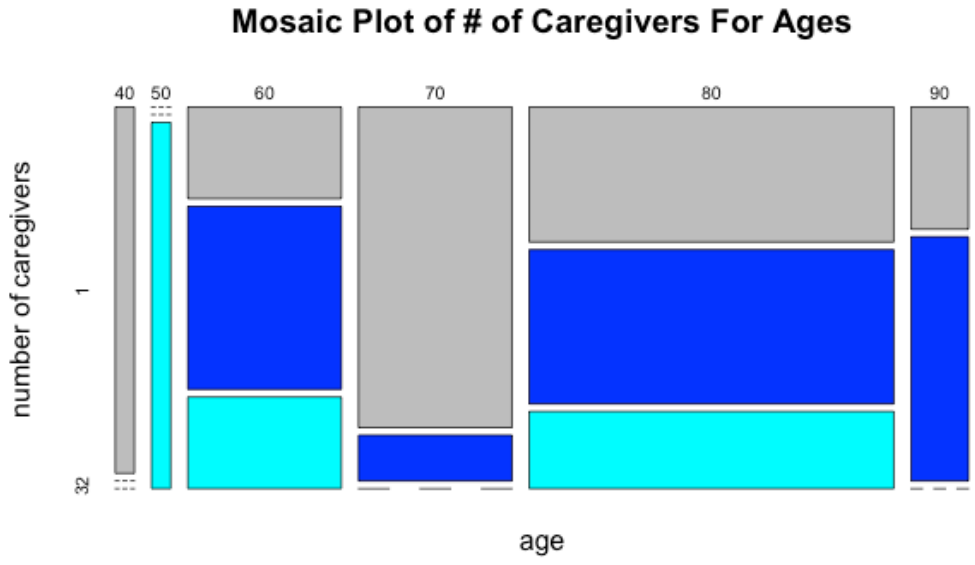


Figure 10 displays the relationship between age and the number of caregivers. Gray bars represent the group of elders with 1 caregiver, while blue bars represent the group of elders with 2 caregivers and mint bars represent the group of elders with 3 caregivers. The area of each bar reveals the amount of elders in a specific age group with a specific number of caregivers.
Out of 137 elders in Third Street Alliance, only 40 individuals have recorded information of caregivers. The reason for 40s and 50s to have only one color of bars is that there is only one person aged 40s and only one person aged 50. Thus, the cases of 40s, 50s, and 90s are special that we cannot get any useful information from the data to make a conclusion about the relationship between age and the number of caregiver since the size is too small. What we could see from the plot is that any number of caregivers (1-3) will be appropriate for 60s and 80s since areas are almost split evenly. For 70s, 1 caregiver will be enough for most elders, but there might be something else influencing the number of caregivers, like the opinions of their family members.

While only 1 caregiver is normally distributed, the distributions of 2 and 3 caregivers are negatively skewed that their medium ages are a little bit higher than the medium age of elders with only 1 caregiver. The range of age of elders with 1 caregiver is the widest since there are two outliers that the minimum age is 45.26 and the maximum age is 97.88, but its spread is the most narrow. It is surprising to see that half of elders who have 3 caregivers are aged in such a narrow interval between 80 and 85. But overall speaking, the samples we have do not show any obvious evidence of age affecting the number of caregivers.

Test for hypothesis

We view male and female as two populations (Nm, Nf) for the number of caregivers and we believe that there is a difference due to the gender. We still performed a two-sided test for them. However, based on the test statistics we get, we cannot conclude that there is a difference in the number of caregivers between women and men. The 95% CI suggest that the difference in the number of caregivers between women and men will be between -0.9279 and 0.5556.

```
t = -0.573, df = 8.4856, p-value = 0.5815
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9278932  0.5555988
sample estimates:
mean of x mean of y
 1.571429  1.757576
```

Conclusion:
According to the t tests we performed, there is no statistical evidence suggest that gender makes the care of the elders different, but situations could be different for male and female if we look at something else besides the services they received.
The statistic shows that age affects the length of services since older men stay shorter as their ages increase while older women stay longer as their ages increase. However, their numbers of caregivers and days per week attend are not significantly different according to the boxplots we made. We consider the elders' family as the actual determinant of those services instead of their ages.