The Impact of Crime Rate on Residential Location Choice in Seattle, WA: an Observational Study

Wasinee Opal Sriapha - Liangzi Zhang - Shivam Jakhanwal
Northeastern University, IE 6200 - Engineering Probability and Statistics, November 2019

Abstract

Background: Selecting your next residential location is a not a decision that should be made lightly. There are many aspects about where you reside that can impact the level of your safety, satisfaction and overall quality of life. The factors that influence home buying are of interest to urban planner and real estate professionals, as well as the general public. Number of studies have been conducted in a few metropolitan areas to assess the factors home buyers/renters consider when deciding on a place to live. This study focuses on Seattle, the city with an ever-growing net in-migration rate and increasing homelessness, addiction, and crime. Do Seattle residents pay attention to the crime rate before moving into a new neighborhood? If so, to what extent? Does gender matter when considering the crime rate in the area? This study is designed to answer these questions, as well as assess how other factors such as commute time to work/school, and proximity of family/friends influences Seattle population of various genders and age groups in deciding where to live.

Objectives: *Primary:* To assess if crime rate is equally prioritized between men and women in Seattle. *Secondary:* To evaluate how crime rate influences Seattle population of various races and age groups in deciding where to live; To evaluate how commute time to work or school influences Seattle population of various age groups in deciding where to live; To evaluate how proximity of family and friends influences Seattle population of various age groups in deciding where to live.

Methods: Fifty-two survey participants who currently reside or have resided in Seattle were randomly selected to complete the study's survey, which asked them to rank levels of importance among crime rate, proximity of family and friends, and commute time to work/school. The study utilized the stratified sampling technique to enable drawing of a sample that would be statistically representative of the Seattle's racial composition. The two-sample t-test was used to test the hypothesis that there would be no significant difference between the true population mean of the crime rate scores ranked by men and the mean of the crime rate scores ranked by women. Other graphical techniques were also used to interpret how proximity of family and friends and commute time influence Seattle population of various genders and age groups in deciding where to live.

Results: The calculated P-value of approximately 0.58 is greater than the significance level of 0.05, hence the test failed to reject the null hypothesis. As a result, there is not sufficient evidence to conclude that the difference between the population means is statistically significant.

Conclusion: There is no difference between men and women regarding crime rate as a factor in deciding the area to reside.

1. Introduction

U.S. Department of Housing and Urban Development published a comprehensive housing analysis paper in May 2017 showing over 3 percent of annual job growth in Seattle from 2014 through 2016. The high growth rate helped Seattle earn its spot on Forbes' 2017 list of America's Fastest-Growing Cities. Between 2010 and 2017, the number of homeowners in Seattle raised to an average of 1.0 percent increase, or 6,600 households annually [1].

Although Seattle's violent crime rates has always been below the national rates, it was our interest to find out how much its current residents considered crime rate before moving into the new area. First, let's take a look at the research that was done in another city with a high concentration of high-technology companies and thus growing immigration rate. Bishop and Murphy published in their literature that the average household in the Bay Area of California was willing to pay $472 per year for a ten percent reduction in violent crime. The research analyzed data between 1990 and 2008. They also discovered that White families had the greatest aversion for crime, while Hispanic families were least to prioritize crime rate. In fact, white families were inclined to spend $4.32 more than Hispanic families to avoid one additional crime per 100,000 residents [2]. What about Seattle? The city is known as one of the multicultural cities of the United States with less than 70% of its residents identifying themselves as white. In this study, we aim to assess the impact of crime rate on residential location choice in Seattle residents of various races, genders, and age groups. The influence of other factors such as proximity of family and friends, and commute time to work/school are also discussed.

2. Study Design and Sampling

2.1 Stratified Sampling

Because the census surveying on the entire Seattle population was a nonviable method in this study, Stratified sampling technique was adopted to enable drawing of a sample that would be statistically representative of the Seattle population. With stratified sampling, the variability within the subgroups is lower compared to the variations that comes with utilizing simple random sampling technique. Due to the high statistical precision that stratified sampling offers, a smaller sample size is required.

To ensure each possible sample is equally likely to occur, *proportionate stratified method* can be applied. This makes the sample size of each racial subgroup to be proportionate to the population size of the stratum. Based on the sample surveyed, the conclusions can be drawn while ensuring that none of the population's segment is underrepresented or overrepresented.

Referencing data on the racial composition of Seattle from US Census website, the population of Seattle was divided into smaller subgroups, or strata, according to demographics. A master list of 67 Seattle residents was created, each person with a unique numerical index. The sampling fraction of 0.773 was given to all strata and a random sample (probability sample) from each stratum was drawn. Using R, 52 survey participants were randomly drawn from the master list according to Seattle's racial proportions. The survey was sent to each participant electronically.

Table 1: Demographics of Survey Participants

Racial Categories	Seattle's Racial Proportion	Number of Identified Potential Participants on Master List	Number of Surveys Conducted (0.8 sampling factor)
White	68.6%	44	34
Asian	14.5%	10	8
Black or African American	7.1%	5	4
Hispanic or Latino	6.5%	5	4
Others*	3.3%	3	2

*American Indian and Alaskan Native, Native Hawaiian and Other Pacific Islander, mixed race

Table 2: Genders of Survey Participants

Genders	Number of Participants
Females	27
Males	25

Table 3: Age Groups of Survey Participants

Age Groups	Number of Participants
18 – 34 years old	25
35 – 50 years old	14
51 - 80 years old	13

2.1.1 Shortcomings of the Collection Method

Since stratified sampling technique was selected for this study to ensure that none of the Seattle population's segment is underrepresented or overrepresented, a specific number of completed surveys were determined for each race. Due to time restrictions of the project and specification of racial data needed for the study, finding respondents from online research communities, such as SurveyCircle was not feasible. The master list used for study participant randomization was compiled with names of residents from researchers' circles of friends. This introduced a bias similar to the one of purposive sampling. The potential survey respondents in each stratum were identified based on a particular characteristic of interest. In this case, the particular characteristic is the type of race. The nature of this bias can make the representativeness of the sample questionable, since it is difficult to prove that the judgement that was used to select potential survey respondents was appropriate.

3. Results

3.1 Exploratory Analysis

This section focuses on exploring the study's secondary objectives. The following are observed:

- How crime rate influences Seattle population of various age groups and races in deciding where to live.
- How commute time to work or school influences Seattle population of various age groups in deciding where to live.
- How proximity of family and friends influences Seattle population of various age group in deciding where to live.

3.1.1 Crime Rate

In order to evaluate how crime rate influences Seattle population of various age groups and races in deciding where to live, Boxplot was used as a visualization tool to interpret the collected data.

Figure 1 shows the Crime Rate Scores of different age groups. The 18 - 32 years old group with a sample size of 25 has the approximate median Crime Rate Score of 3. As shown in the box plot, the median line of this age group is closer to the bottom quartile, thus the box is positively skewed. This implies that the data set contains higher frequency of high rating scores for the crime rate factor. Two of the later age groups, the 35 - 50 years old group (sample size of 14) and the 51 - 80 years old group (sample size of 13) have the approximate median Crime Rate Score of 4. The upper whisk is also not shown above the upper quartile for both age groups. This suggests that the upper quartile is equal to the maximum, since the whiskers indicate data's minimum and maximum. An outlier is also noted for the 51 - 80 years old group. This indicates that there is a data point lower than the 1.5*IQR below the first/lower quartile.

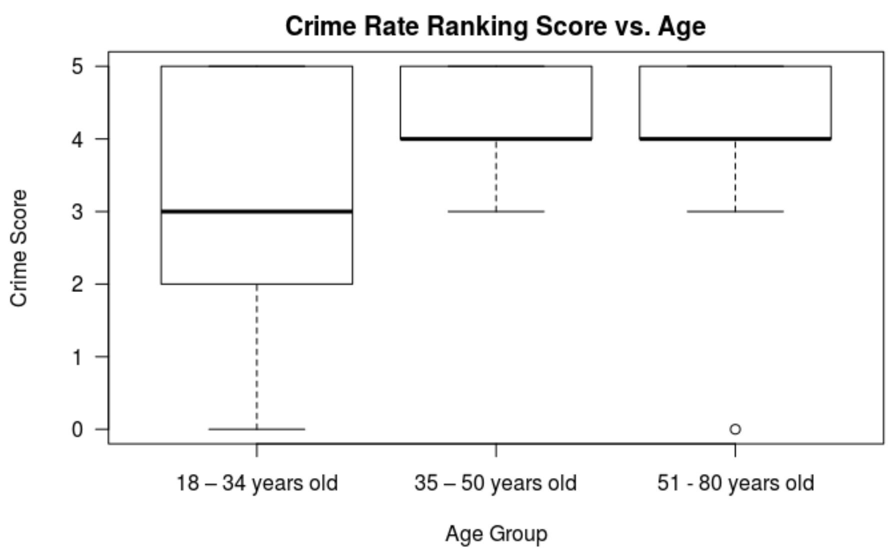

Figure 1: Boxplot of Crime Rate Score vs. Age Groups

The relationship between the Crime Rate scores and different races can be seen in *Figure 2*. Asian has the highest mean score of over 4.125 when it comes to aversion to high crime areas. White, Black and Hispanic have the mean Crime Rate Score in the range of 3.5 to 4, while the lowest mean score of approximately 2 is from the participant group who categorized themselves in the "Others" race group.

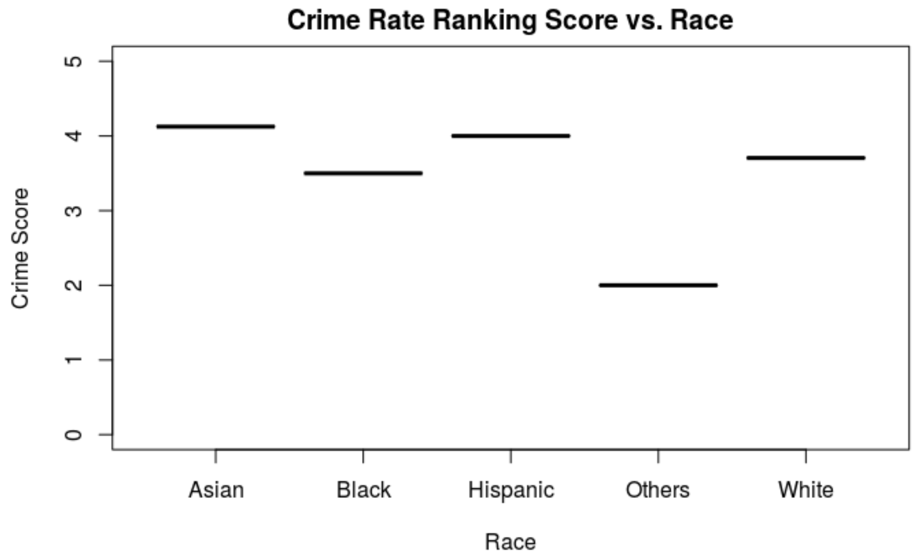

Figure 2: Plot of Mean Crime Rate Score vs. Race

3.1.2 Commute Time to Work or School

The relationship between the Commute Time to Work or School Score (CTWS) and the different age groups are shown in *Figures 3.* The 18 - 32 years old group with a sample size of 25 has the approximate CTWS of 4. There is an outlier of 2 present in this age group implying that its value is lower than the 1.5*IQR below the first/bottom quartile. The 35 - 50 years old group with the sample size of 14 has the approximate median CTWS of 3 and a symmetric distribution. In other words, the box has equal proportions around the median. The lower whisker is also not present under the bottom quartile suggesting that the bottom quartile is equal to the minimum. The 51 - 80 years old group with the sample size of 13 has the approximate median CTWS of 4. Even though the box has the lower fence line of 0, it is negatively skewed and displays the median line closer to the top quartile. This implies that the data set of this age group has a higher frequency of low valued scores.

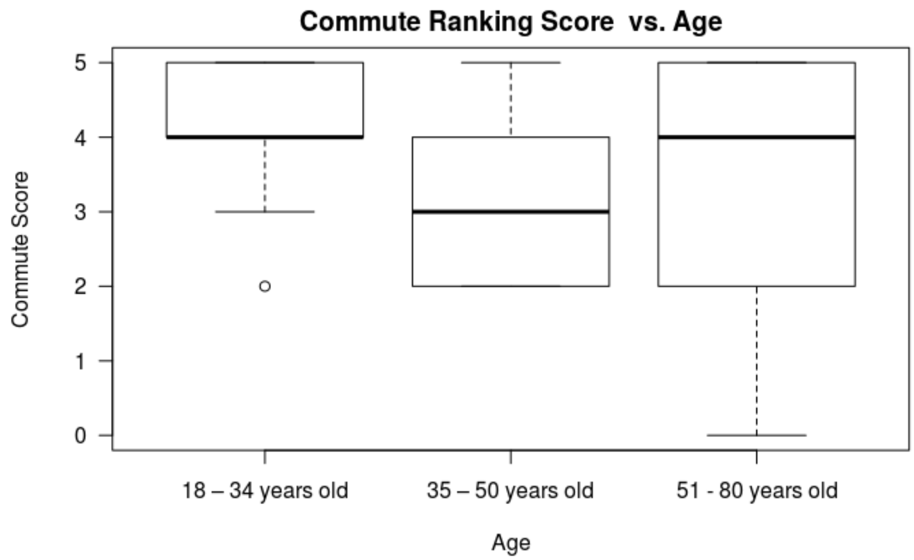

Figure 3: Boxplot of Commute Time to Work or School Score vs. Age Groups

3.1.3 Proximity of Family and Friends

The relationship between the Proximity of Family and Friends Score (PFFS) and the different age groups are shown in *Figures 4.* The 18 - 32 years old group with a sample size of 25 has the approximate median PFFS of 3. Moreover, two outliers of 0 and 1 are present in this age group. This implies that the values of the two outliers are lower than the 1.5*IQR below the first/lower quartile. The 35 - 50 years old group with the sample size of 14 has the approximate median PFFS of 3. Lastly, the 51 - 80 years old group with the sample size of 13 has the approximate median PFFS of 4. Since the upper quartile of this age group is equal to the maximum, upper whisk is not shown above the upper quartile. It has also been noted that the box for the 35 - 50 years old group and the one for the 51 - 80 years old group have symmetric distribution.

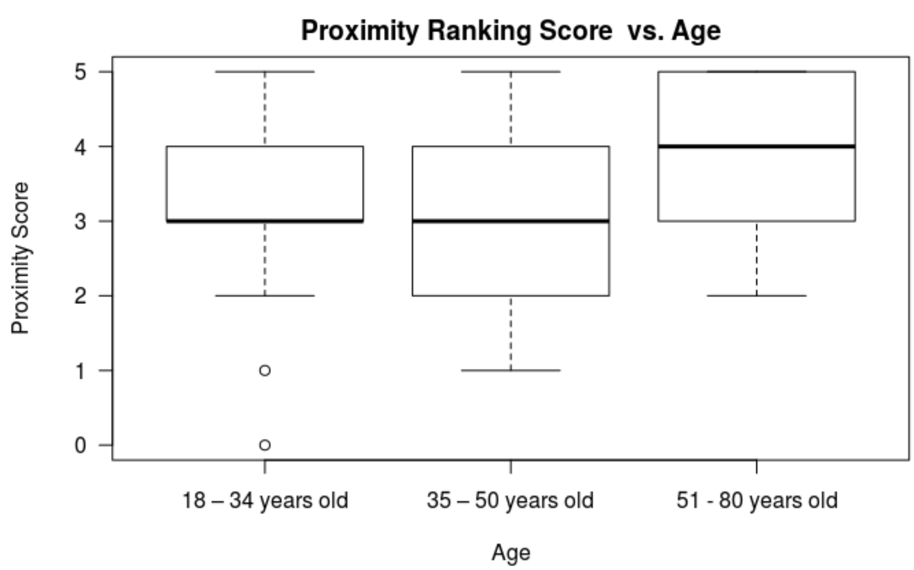

Figure 4: Boxplot of Proximity of Family and Friends Score vs. Age Groups

3.2 Statistical Analysis

3.2.1 Q-Q Plot

To check whether the population data is normally distributed, the collected data is put in its own quantile and plotted against its theoretical quantiles (theoretical normal distribution). If both sets of quantiles are from the same distribution, the plotted points should form a line that is roughly straight. In this study, the qqnorm() function will be utilized to plot the quantile from a standard Normal distribution against the quantile from the Crime Rate Score for all genders, the quantile from Male Sample, the quantile from Female Sample.

Moreover, the quantiles of the data can be plotted against one another to test if they are identical in the two samples. The qqplot () function will be utilized to draw a conclusion on whether the Crime Rate scores of the two samples are distributed equally.

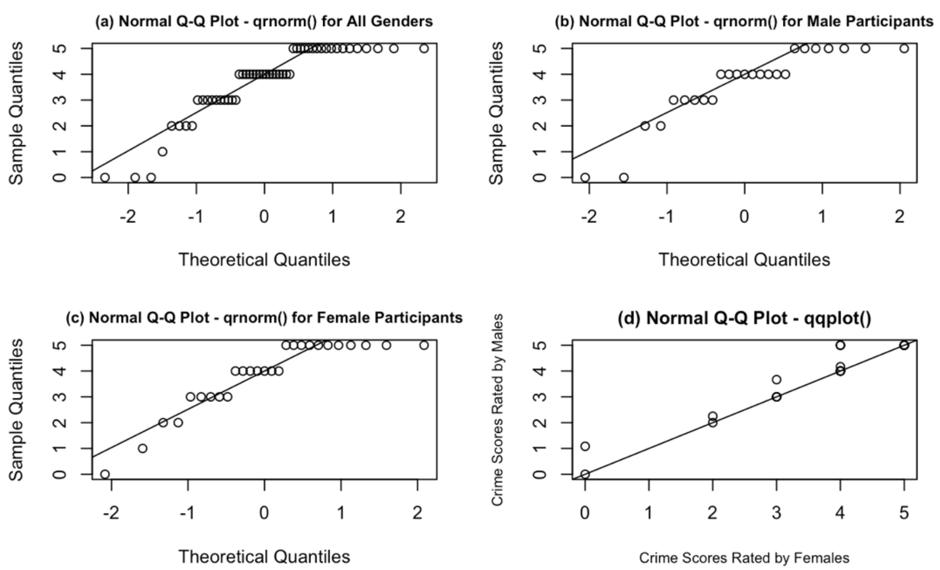

Figure 5: Q-Q Plots

Fig. 5(a) The quantile of the Crime Rate Score for all genders was plotted against the quantile from a standard Normal distribution.
Fig. 5(b) The quantile of the Crime Rate Score for Male sample was plotted against the quantile from a standard Normal distribution.
Fig. 5(c) The quantile of the Crime Rate Score for Female sample was plotted against the quantile from a standard Normal distribution.
Fig. 5(d) When quantiles of Male sample and Female sample are plotted against one another, it can be noted that most of the points on the plot fall on the 45-degree line. This suggests that the distributions of are mostly identical. Some of the data points that do not fall on the line suggest a few more-extreme values on either end of the distribution.

3.2.2 Two-Sample t-Test

The two-sample t-test is used to test the hypothesis that the population means are equal for the two samples. The assumption that the variances for the two samples are equal is also made.

In order to examine whether gender factor influences how crime rate is prioritized when Seattle residents decide where to live, this analysis method was performed.

Variables: Gender is a categorical variable creating two independent samples, male and female. The numeric outcome variable is the rank score of crime rate as a factor in deciding where to live.

Hypothesis: Theoretically, two-sample t-test for unpaired data is defined as:

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$

The null hypothesis, H_0 in this study is defined as:
"There is no significant difference between the true population mean of the crime rate scores ranked by men and the mean of the crime rate scores ranked by women"

The alternative hypothesis, H_a in this study is defined as:
"There is indeed a significant difference between the true population mean of the crime rate scores ranked by men and the mean of the crime rate scores ranked by women."

Test statistic (t) is defined by the following equation:

$$T = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{S_1^2/N_1 + S_2^2/N_2}} \quad ,$$

where N_1 and N_2 are the sample size of male and the sample size of female respectively, $\overline{Y}_1$ and $\overline{Y}_2$ are the sample means of male and the sample means of female respectively, and S_1^2 and S_2^2 are the sample variances of male and the sample variances of female respectively.

P-Value:
The analysis involves comparing the P-value to the significance level (0.05 will be used for this study) and rejecting the null hypothesis if the P-value is less than the significance level.

Two approaches of calculating the p-value were carried out:

By using the t-test statistical method in R, the p value was calculated and returned 0.5823635. The lower bound of the confidence interval is -0.5804736 and the upper bound of the confidence interval is 1.010103.

R Built In t-test function, t.test ( ), was also utilized to confirm the accuracy of the result. The confidence interval was set to 95 percent giving the significance level (α) of 0.05.

Below is the returned result:

Welch Two Sample t-test
data: mydata$Crime[mydata$Gender == "Female"]
and mydata$Crime[mydata$Gender == "Male"]
t = 0.55748, df = 49.317, p-value = 0.5797
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.5594151 0.9890447
sample estimates:
mean of x mean of y
3.814815 3.600000

The p-value from this test is 0.5797, the lower bound of the confidence interval is -0.5594, the upper bound of the confidence interval is 0.9890.

Table 4: The comparison between R built in T-Test and the t-test statistical method

	R Built in T-Test	Using Test Statistic
P-value	0.5797	0.5824
Lower bound of CI	-0.5594	-0.5805
Upper bound of CI	0.9890	1.0101

As shown in the table above, the differences between the results from the two methods are relatively small.

Interpretation:

The P-value of approximately 0.58 is greater than the significance level of 0.05, hence the test failed to reject the null hypothesis. As a result, there is not sufficient evidence to conclude that the difference between the population means is statistically significant. In other words, the null hypothesis that "*there is no significant difference between the true population mean of the crime rate scores ranked by men and the mean of the crime rate scores ranked by women*" was not rejected at the $\alpha = 0.05$ level. With 95% confidence, the difference between the two true population means is likely to be between the lower limit of -0.5805 and the upper limit of 1.0101. The null hypothesized difference of 0 falls in the confidence interval, therefore confirming the failure to reject the null hypothesis.

Sampling Distribution:

It is inappropriate to use a small data set to make theoretical generalizations to the studied population. For example, data gathered from only 10 survey participants should not be used to assess how crime rate influences Seattle residents when choosing a place to reside. The sample size would be too small to be used to make inference to the entire Seattle population. As the sample size increases, the mean of the sampling distribution will approach the population mean.

This is referring to the Central Limit Theorem, which is essential for computing the normal probability when the sample means are used to draw conclusions about a population mean. As the observation in this study was repeated 1000 times with the sample size of 30, it was expected that the resulted distribution of sample means will be approximately normal – as shown in the *Figure 6.*

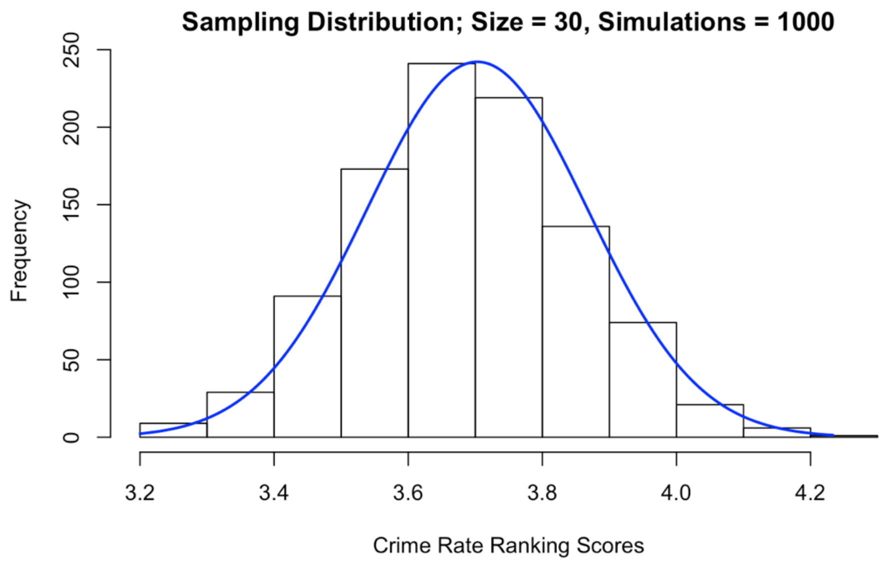

Figure 6: Sampling Distribution

4. Discussion

4.1 Exploratory Analysis

4.1.1 Crime Rate

The results suggest that middle-aged adults (ages 36-55 years) and older adults (aged 51 - 80 years) in Seattle take crime rate in the area into serious consideration when they are deciding where to live. This is shown in the Boxplot where the box's top quartile is equal to the maximum and the lower fence value is 3. The median Crime Rate Score of the two groups is also as high as 4. The range of values on the same factor is wide for the young adults (ages 18-34 years), even though the top quartile also reaches the maximum. This indicates that most young adults do consider crime rate before relocation, while a few of them do not.

In addition, this study conveys that Asian households in Seattle had the greatest aversion for crime. This is contrasting the research conducted by Bishop and Murphy in the Bay Area of California where White households prioritized crime rate more than any other races. There is no significant difference among White, Black and Hispanic ratings. Interestingly, the participant group who categorized themselves in the "Others" race group seemed to be least concerned about crime rate.

4.1.2 Commute Time to Work or School

Both young adults (ages 18-34 years) and older adults (aged 51 - 80 years) has the upper quartile in the Boxplot equivalent to the maximum. The assumption can be made that the young adults are mostly students or employees. It is likely that they prefer to live closer to their schools or offices. Some of the young adults might not be car owners, therefore a long-distance commute might not be feasible. The Boxplot for older adults has the lower fence line of zero instead of some low value outliers. This suggests that some adults in Seattle couldn't care less about the commute time. The reason for this might be that they are retired and no longer have to commute to work. The data points that land on the top quartile for this age group could have been from the older adults who are still working. The mean and median for middle-aged adults (ages 36-55 years) in this category are neither drastically high nor low. This supports the assumption that most of the middle-aged adults are employees and own a car.

4.1.3 Proximity of Family and Friends

According to the Boxplot, it is obvious that the majority of older adults (aged 51 - 80 years) prefer to live closer to their friends and families. A great number of people in this age range are retired and likely have more spare time to spend with families and friends. This might contribute to them paying more attention to the proximity to loved ones when choosing a place to live. Middle-aged adults (ages 36-55 years), on the other hand, seem to prioritize this category less than the other age groups.

4.2 Statistical Analysis

Since the p value is 0.5824 which is greater than the significant level of 0.05, the conclusion can be made that gender doesn't influence how crime rate is prioritized when Seattle residents decide where to live.

4.3 Implications

As mentioned earlier, this research is a great resource mainly for urban planner and real estate professionals because they may refer to the findings in this research when predicting which factors may influence their Seattle's future customers.

Most importantly, this research proved that crime rate in the area is equally prioritized between men and women of Seattle before they decide where to reside. Since the stratified sampling was used, none of the Seattle's racial segment is underrepresented or overrepresented. The data on how crime rate influences Seattle population of various races may help the real estate professionals work with each customer of a certain race more efficiently. The data analyzed for each age group is also useful in predicting how commute time and proximity of family and friends influence each potential customer of a particular age.

4.4 Limitations

The findings of this study have to be seen in light of some limitations. While the sampling method for the research was carefully reviewed and chosen, the results are limited by time constraints on sample collection. The nature of sampling steps we followed led to the bias that could make the representativeness of each sample in the population questionable. It was difficult to tell whether the judgement that was introduced in order to select the potential survey respondents was appropriate.

Due to the smaller population size of 52, it was also difficult to find significant trends and relationships for some of the strata such as Hispanic or Latino and Others (American Indian and Alaskan Native, Native Hawaiian and Other Pacific Islander, mixed race). Larger sample size will confirm a representative distribution of the population and can be considered a solid representative of groups that the results are meant to generalize.

4.5 Next Step

The limitations of this research suggest topics to be investigated in the future. The following are a few of them:

- Since the people who responded to this research's questions might have not truly been a random sample, the bias described in the Shortcomings of the Collection Method section was unavoidable. The future studies can examine whether the theoretical generalizations made from the studied sample will be different if this bias is removed.
- Future studies with a larger population size can better assess how the studied factors, such as crime rate, influence some of the smaller racial segments in Seattle.

Appendix

i. References

[1] Comprehensive Housing Market Analysis for Seattle-Bellevue-Everett, Washington Housing Market Area. (2017). *U.S. Department of Housing and Urban Development* . Retrieved from https://www.huduser.gov/portal/publications/pdf/SeattleWA-comp-17.pdf

[2] Bishop, K. C., & Murphy, A. D. (2011). Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach. *American Economic Review*, *101*(3), 625–629. doi: 10.1257/aer.101.3.625

ii. Statistical Computing in R

```
require("readxl")
library(dplyr)
require(mosaic)
mydata <- read_excel("ZOS.xlsx")
attach(mydata)
str(mydata)
summary(mydata)

# Including Plots

# Crime Rate vs. Age

boxplot( mydata$Crime ~ mydata$Age,
        notch = F,
        las = 1,
        ylab = "Crime Score",
        xlab = "Age Group",
        main= "Crime Rate Ranking Score vs. Age")

# Crime Rate vs. Race

meanWhite <- mean(mydata$Crime[mydata$Race=='White'])
meanAsian <- mean(mydata$Crime[mydata$Race=='Asian'])
meanOthers <- mean(mydata$Crime[mydata$Race=='Others'])
meanHispanic <- mean(mydata$Crime[mydata$Race=='Hispanic or Latino'])
meanBlack <- mean(mydata$Crime[mydata$Race=='Black or African American'])

mean1 <- c(meanAsian,meanWhite, meanHispanic, meanOthers,meanBlack)
race1 <- c("Asian","White","Hispanic","Others","Black")

df1 <- data.frame(race1,mean1)
plot(df1, ylim = c(0,5), xlab="Race",ylab="Crime Score", main = "Crime Rate Ranking Score
vs. Race")
```

```
# Proximity Score vs. Age

boxplot( mydata$Proximity ~ mydata$Age,
        notch = F,
        las = 1,
        ylab = "Proximity Score",
        xlab = "Age",
        main= "Proximity Ranking Score  vs. Age")


# Commute Score vs. Age

boxplot( mydata$Commute ~ mydata$Age,
        notch = F,
        las = 1,
        ylab = "Commute Score",
        xlab = "Age",
        main= "Commute Ranking Score  vs. Age")

# Q-Q Plot

par(mfrow=c(2,2))

qqnorm(mydata$Crime,  main='(a) Normal Q-Q Plot - qrnorm() for All Genders', cex.main=0.85
)
qqline(mydata$Crime)

qqnorm(mydata$Crime[mydata$Gender=="Male"],  main='(b) Normal Q-Q Plot - qrnorm() for
Male Participants', cex.main=0.85)
qqline(mydata$Crime[mydata$Gender=="Male"])

qqnorm(mydata$Crime[mydata$Gender=="Female"],  main='(c) Normal Q-Q Plot - qrnorm() for
Female Participants', cex.main=0.85)
qqline(mydata$Crime[mydata$Gender=="Female"])

qqplot(mydata$Crime[mydata$Gender=="Male"], mydata$Crime[mydata$Gender=="Female"],
main = "(d) Normal Q-Q Plot - qqplot()",
      xlab = "Crime Scores Rated by Females", ylab = "Crime Scores Rated by Males",
cex.main=1.00, cex.lab=0.8, cex.axis=1)
abline(0,1)

# Two Sided T-Test

# Calculation of P value

mydata <- read_excel("ZOS.xlsx")
# the parts of the test statistic
```

```
# sample means
x_bar_f <- mean(mydata$Crime[mydata$Gender=="Female"])
x_bar_m <- mean(mydata$Crime[mydata$Gender=="Male"])
# null hypothesized population mean difference between the two groups
mu_0 <- 0
# sample variances
s_f_sq <- sd(mydata$Crime[mydata$Gender=="Female"])**2
s_m_sq <- sd(mydata$Crime[mydata$Gender=="Male"])**2
# sample size
n_f <- length(mydata$Crime[mydata$Gender=="Female"])
n_m <- length(mydata$Crime[mydata$Gender=="Male"])
# t-test test statistic
t <- (x_bar_f - x_bar_m - mu_0)/sqrt((s_f_sq/n_f) + (s_m_sq/n_m))
# one sided upper p-value
pval <- pt(q = t, df = min(n_f, n_m)-1, lower.tail = FALSE)*2
pval

# Calculation Of Confidence Interval

#lower bound
L_bound <-(x_bar_f-x_bar_m)+(qt(0.025, min(n_f, n_m)-1)*sqrt((s_f_sq/n_f) + (s_m_sq/n_m)))
L_bound

#upper bound
U_bound <-(x_bar_f-x_bar_m)+(qt(0.975, min(n_f, n_m)-1)*sqrt((s_f_sq/n_f) +
(s_m_sq/n_m)))
U_bound

# R Built In T-Test

t.test(mydata$Crime[mydata$Gender=="Female"],mydata$Crime[mydata$Gender=="Male"])


# Sampling Distribution

set.seed(110)
X_bar10 <- do(1000) * mean(sample(mydata$Crime,30))
h3<- hist(X_bar10$mean,main = "Sampling Distribution; Size = 30, Simulations = 1000", xlab =
'Crime Rate Ranking Scores')
xfit<-seq(min(X_bar10$mean),max(X_bar10$mean),length=250)
yfit<-dnorm(xfit,mean=mean(X_bar10$mean),sd=sd(X_bar10$mean))
yfit <- yfit*diff(h3$mids[1:2])*length(X_bar10$mean)
lines(xfit, yfit, col="blue", lwd=2)
```

iii. Raw Data Set

IP Address	Name	Age Group	Gender	Race	Crime	Commute	Proximity
174.216.21.136	Aimee	35 – 50 years old	Female	White	3	3	3
75.172.122.152	Cherise	18 – 34 years old	Female	White	2	4	4
67.170.62.204	Vicki	51 - 80 years old	Female	White	5	5	5
67.170.60.87	Linda S	51 - 80 years old	Female	White	3	2	5
75.172.188.11	Brian	51 - 80 years old	Male	White	4	5	4
107.77.205.90	Linda	51 - 80 years old	Female	White	4	5	2
50.46.213.96	Camilla	18 – 34 years old	Female	White	5	5	0
67.100.127.68	Jsb	51 - 80 years old	Male	White	3	4	4
67.100.127.68	Linda	51 - 80 years old	Female	White	4	5	2
67.170.60.87	Jim S	51 - 80 years old	Male	White	5	2	4
67.182.145.82	Kevin	18 – 34 years old	Male	Asian	4	4	1
67.182.145.82	Yiwei	35 – 50 years old	Female	Asian	5	2	3
73.35.141.88	Skylar	35 – 50 years old	Female	White	4	5	2
67.182.145.82	Nick	35 – 50 years old	Male	Asian	4	3	1
73.35.141.88	Lorenzo	35 – 50 years old	Male	Hispanic	5	3	5
172.58.45.188	M	51 - 80 years old	Male	Others	0	0	5
184.53.33.9	Bethany	18 – 34 years old	Female	White	5	5	3
73.35.251.31	Hayley	18 – 34 years old	Female	White	2	5	3
207.38.162.211	weather	18 – 34 years old	Female	Asian	5	5	5
68.4.69.207	jiahui	18 – 34 years old	Female	Asian	5	5	5
76.121.241.225	Jonathan	18 – 34 years old	Male	White	3	4	4
174.217.13.163	Karly	18 – 34 years old	Female	White	0	3	5
74.113.40.50	Brooke	35 – 50 years old	Female	White	4	5	3
76.121.241.225	Alli	18 – 34 years old	Female	Hispanic	4	3	4
107.77.209.4	ZJD	18 – 34 years old	Male	Asian	5	3	3
154.57.252.222	LEI	18 – 34 years old	Male	Asian	0	2	3
73.35.178.143	Camille	18 – 34 years old	Female	White	1	4	3
216.243.4.227	Malcolm	18 – 34 years old	Male	White	5	4	3
46.244.28.57	Stephanie	18 – 34 years old	Female	White	3	4	4
104.238.46.169	Cory	35 – 50 years old	Male	White	4	3	3
46.244.28.63	Aminata	35 – 50 years old	Female	Black	4	3	5
46.244.28.62	Maren	51 - 80 years old	Female	White	5	1	4
173.239.198.246	Angelica	18 – 34 years old	Female	Black	3	5	4
73.140.247.97	Adam	18 – 34 years old	Male	White	2	5	4
129.10.7.133	Jeremy	35 – 50 years old	Male	White	5	5	2
40.133.239.214	Tiffany S	35 – 50 years old	Female	White	3	4	3
174.224.0.202	Steve K	18 – 34 years old	Male	Black	2	4	2
71.212.117.37	Mei	35 – 50 years old	Male	Others	4	2	4

IP Address	Name	Age Group	Gender	Race	Crime	Commute	Proximity
174.216.22.203	Jake D	18 – 34 years old	Male	White	3	4	3
71.19.252.125	Ken	51 - 80 years old	Male	White	4	2	3
172.58.45.211	Lisa	51 - 80 years old	Female	White	5	5	5
172.58.45.80	Bruce	51 - 80 years old	Male	White	5	4	3
45.41.181.92	Josh	35 – 50 years old	Male	White	4	2	2
205.164.5.243	Gretchen	35 – 50 years old	Female	White	5	2	3
67.183.13.238	JCN	18 – 34 years old	Male	White	4	4	2
129.10.7.131	Kikii	35 – 50 years old	Female	Black	5	4	4
23.252.62.110	Theo	18 – 34 years old	Male	White	5	3	4
96.76.49.133	Jose N	18 – 34 years old	Male	Hispanic	3	4	4
67.170.48.111	Ethan	18 – 34 years old	Male	White	3	3	4
67.161.122.15	Andy	18 – 34 years old	Male	White	4	4	3
73.225.53.110	Pat	18 – 34 years old	Female	Asian	5	4	3
71.231.7.211	Jennifer	51 - 80 years old	Female	Hispanic	4	5	3