Food Pantry Analysis

Introduction

This report gives an overview of Wesley and Parkland Food Pantry visitor trends over previous years. We utilized data from the Wesley Food Pantry starting in 2012, and the Parkland Food Pantry starting in 2014. We also combined the 2 data sets to see how they compared to each other in recent years. We created a total of 5 graphs to provide an overview of the trends we noticed during our initial analysis. We also ran some statistical tests to determine the significance of the trends we visualized.

```
data = read.csv(file = "/Users/omachowda/Downloads/combined.csv")
```

Data

We were given ten different Excel files which contained data for Wesley and Parkland food pantry from 2012 to 2017. Each Excel file had different formatting which made data aggregation slightly more difficult. We created a single Excel file which contained total visitors (Ind) for each pantry location by month and year. The variables for the new Excel file are Year, Month, PFP, WEFP, and Total. We had lots of missing values because the data collection was different for each year and month. For Parkland Pantry, we did not have data for anything before July 2014 so it was difficult to create total trend graph. We also did not have the data from January 2014 to June 2014 for both locations which why some graphs have breaks in them. We had to make sure we got proper values because it was difficult to work with different formatting for the Excel files. We created the dataset for analyzing visitor trends over time.

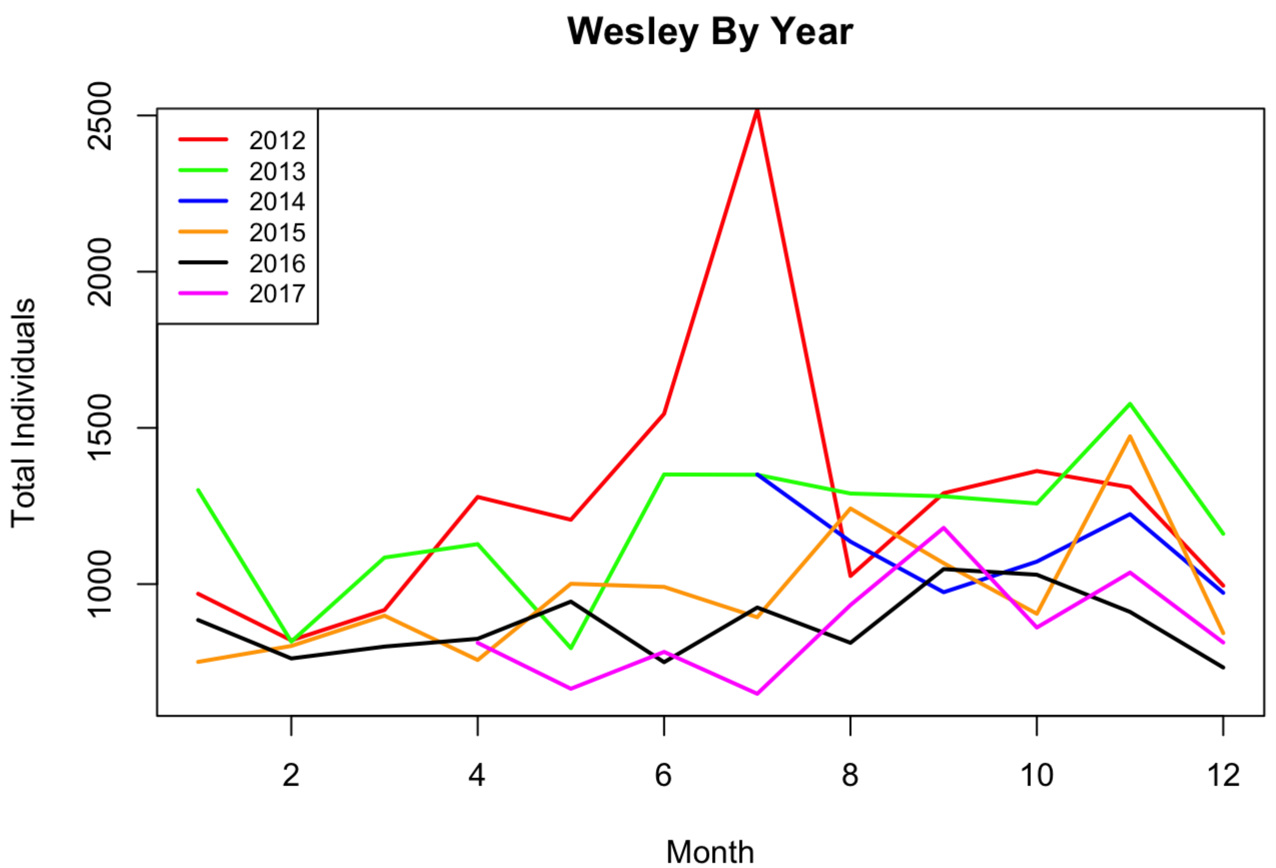
Methods

The graphs were created using the statistical programming language called R. We loaded the newly created CSV file to R. We then split the data by Year to create the graphs. We used the plot function in R to create the different situational graphs. We had to create sequences for the data to generate the lines in the graph so it plots it properly since we had missing data.

Graphs

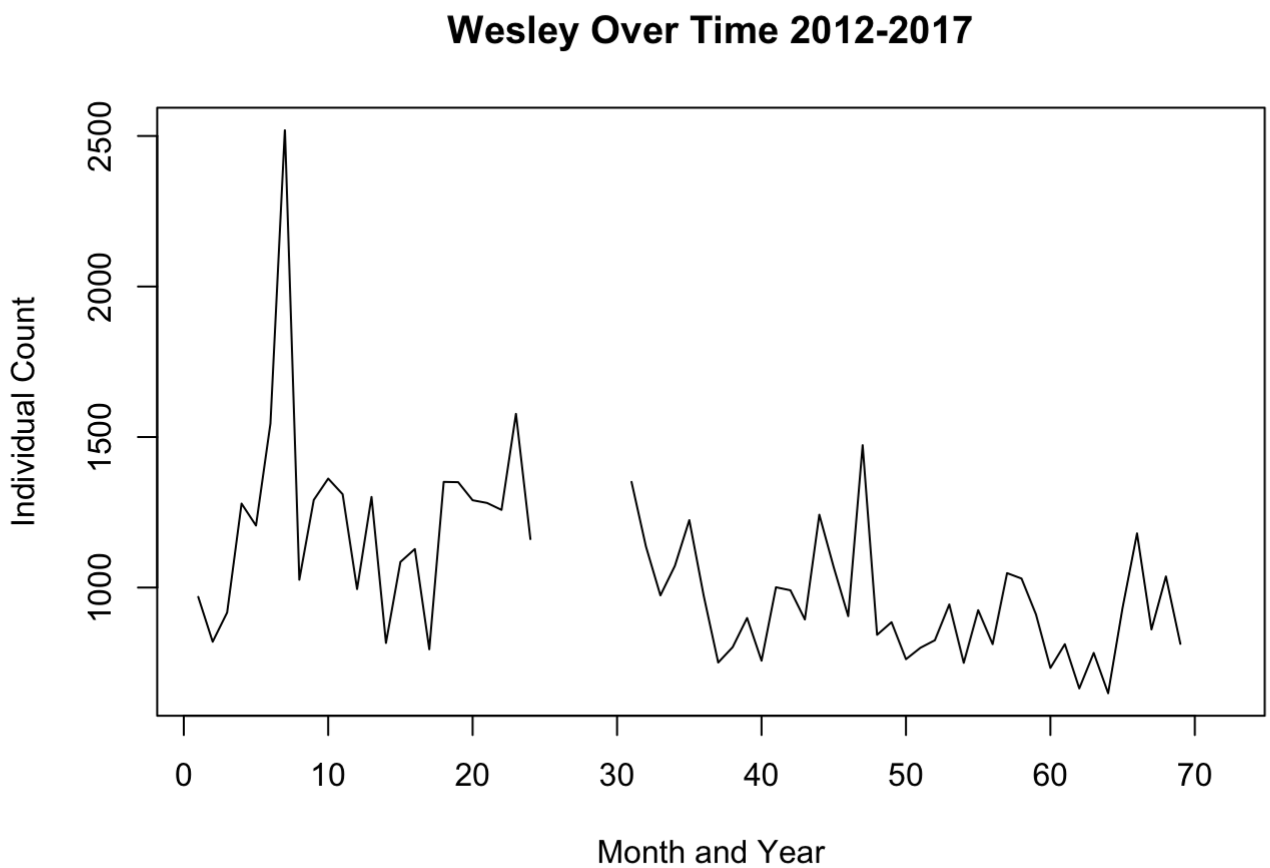
```
#WEFP Plot by Year
splitted = split(data$WEFP, ceiling(seq_along(data$WEFP)/12))

{plot(seq(1,12), splitted$'1',type = "l",lwd = 2,main = "Wesley By Year",ylab = "Total I
ndividuals",col="red",ylim = c(650,2450),xlim = c(1,12), xlab= "Month")
lines(seq(1,12), splitted$'2',type = "l",col = "green",lw = 2)
lines(seq(7,12), splitted$'3'[7:12],type = "l", col = "blue", lw = 2)
lines(seq(1,12), splitted$'4',type = "l", col = "orange",lw = 2)
lines(seq(1,12), splitted$'5',type = "l" , col="black",lw = 2)
lines(seq(4,12), splitted$'6'[1:9],type = "l", col = "magenta",lw = 2)
legend("topleft", c("2012", "2013", "2014","2015","2016","2017"), lty = c(1, 1,1,1,1,1),
lw=2,
col = c("red", "green","blue","orange","black","magenta"), cex = .8)
}
```



From this plot we see that the busiest months for all years seem to be September through December. We suspect this is due to the large number of holidays in this period of time.

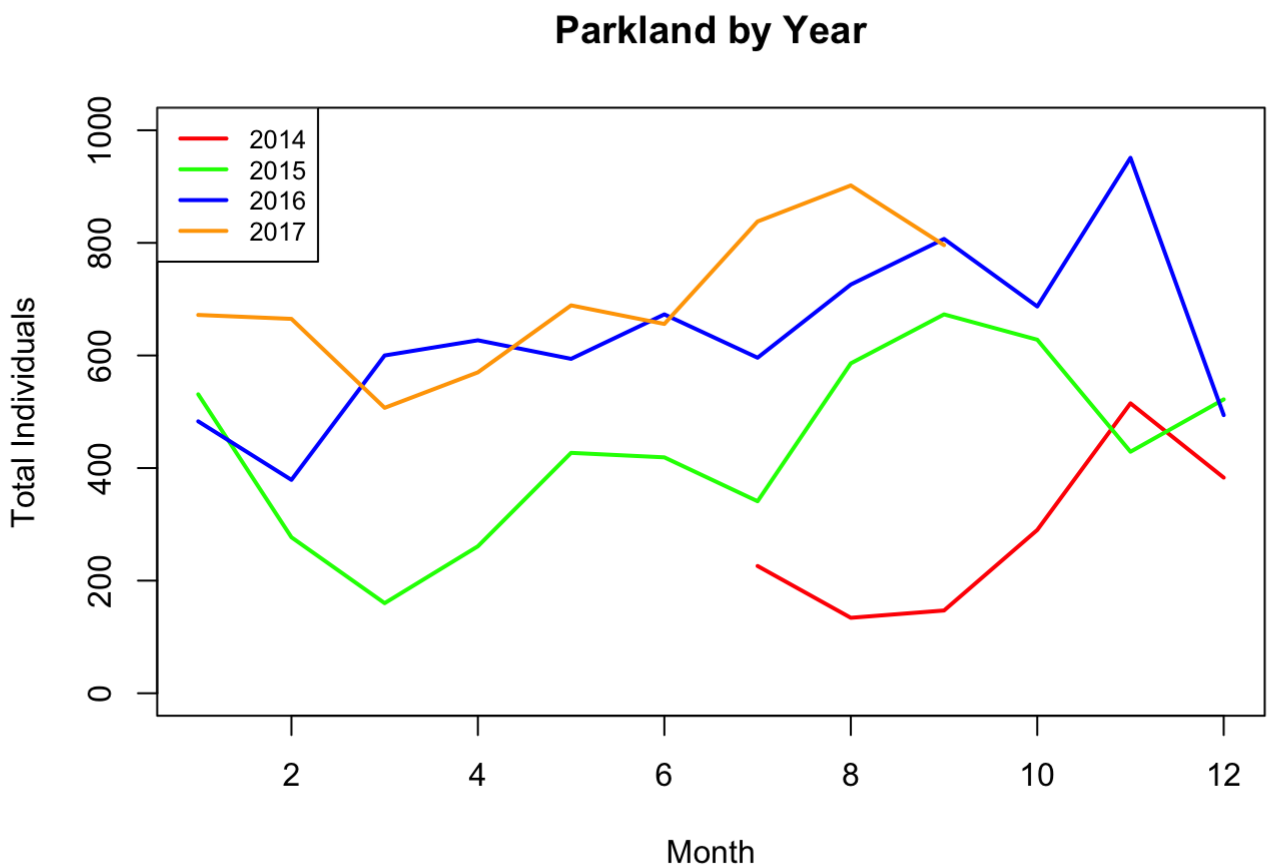
```
#WEFP Overtime
#explain missing data
plot(seq(1,72), data$WEFP, type="l",xlab= "Month and Year",ylab="Individual Count",main=
"Wesley Over Time 2012-2017")
```



With this graph we see the number of people graduaully decrease as time goes on. We think this maybe due to the opening of the Parkland Food Pantry in 2014. We make a compartive graph of this data below.

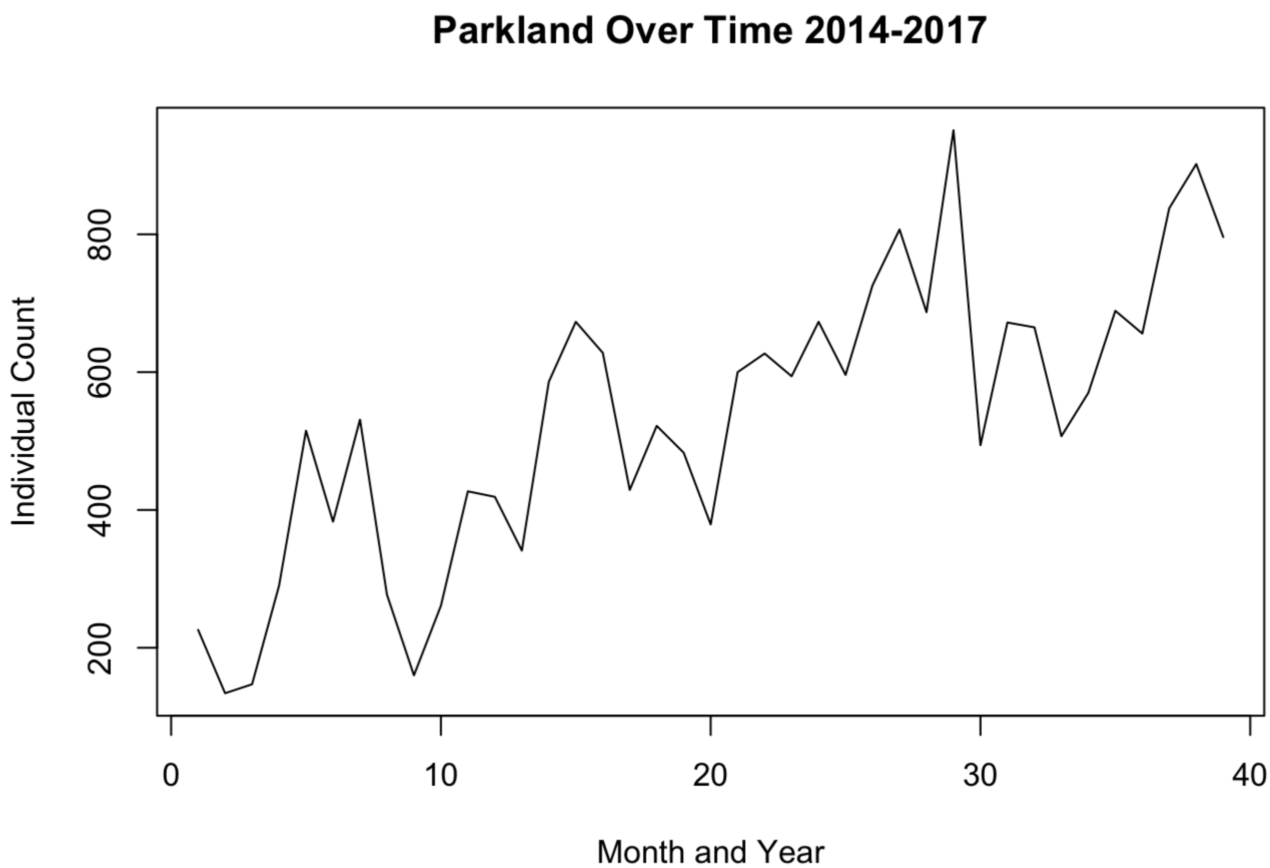
```
#PFP Plot
splitted2= split(data$PFP, ceiling(seq_along(data$PFP)/12))

{plot(seq(1,12), splitted2$'3',type = "l",col="red", ylim = c(0,1000),lwd=2,xlab = "Mont
h",ylab="Total Individuals",main="Parkland by Year")
lines(seq(1,12), splitted2$'4',type = "l",col="green",lw=2)
lines(seq(1,12), splitted2$'5',type = "l",col="blue",lw=2)
lines(seq(1,9), splitted2$'6'[1:9],type = "l",col = "orange",lw=2)
legend("topleft", c("2014","2015","2016","2017"), lty = c(1, 1,1,1),lwd=2,
col = c("red", "green","blue","orange"), cex = .8)
}
```



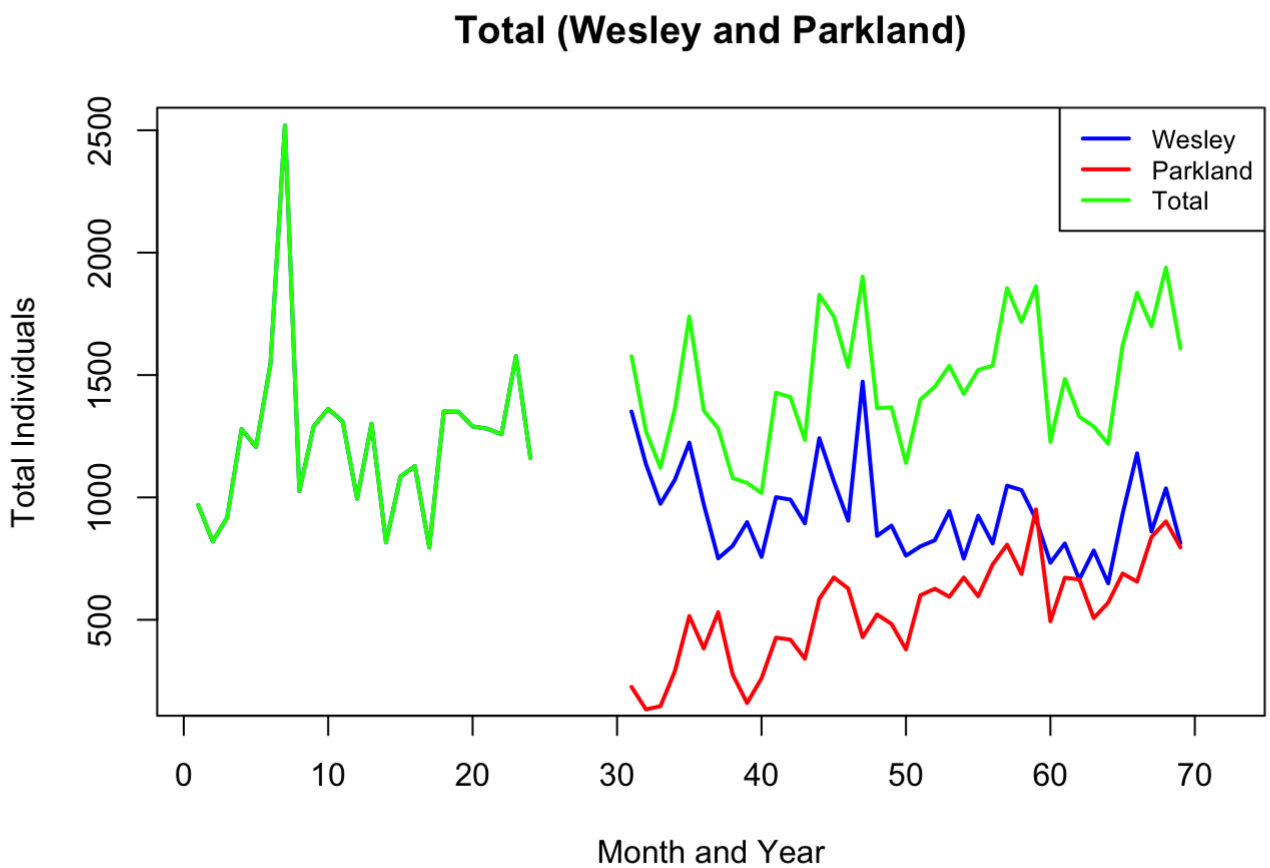
For the Parkland Food Pantry we see that August and November seem to be the busiest months.

```
#PFP Overtime
plot(seq(1,39), data$PFP[31:69], type="l",xlab= "Month and Year",ylab="Individual Count"
,main="Parkland Over Time 2014-2017")
```



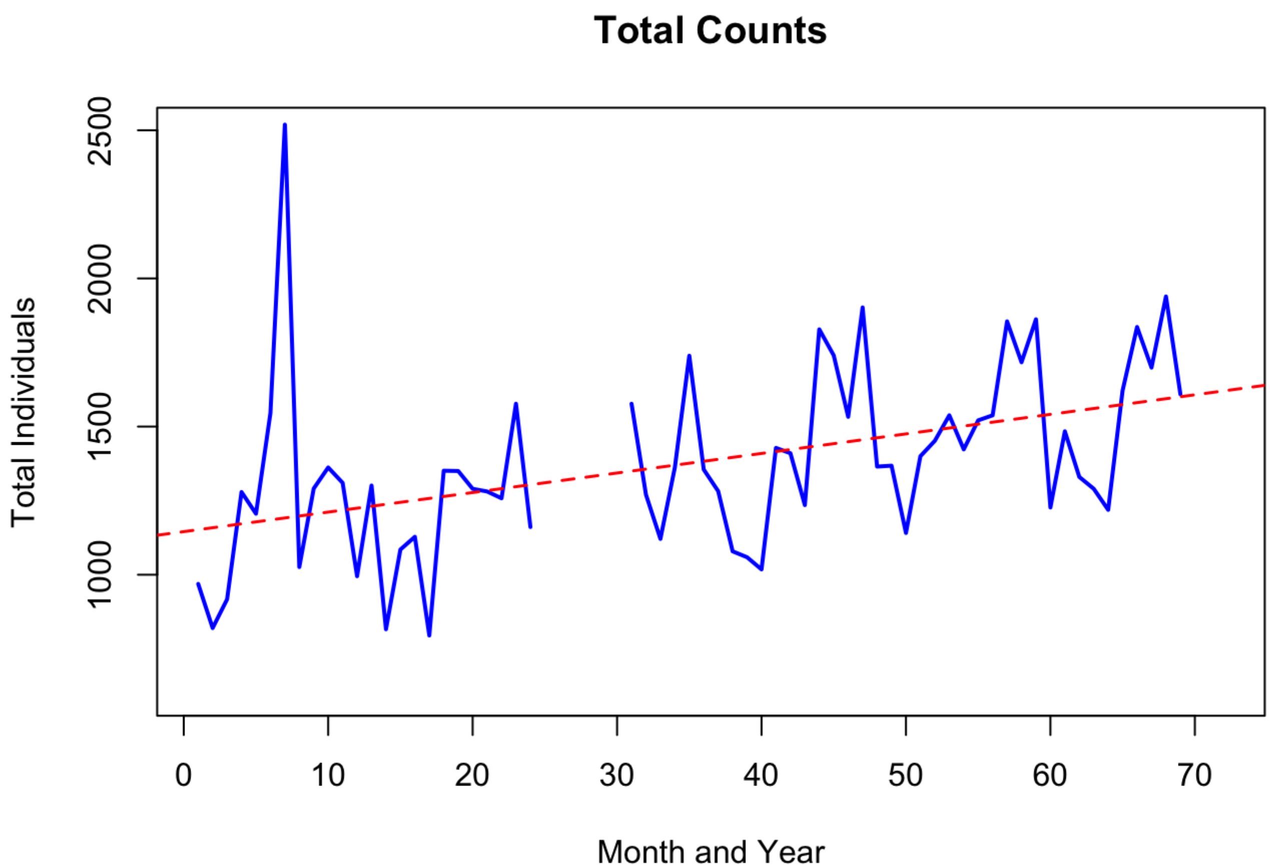
With Parkland we see a definite increase since its opening in 2014. This maybe due to the fact that people located closer to parkland will go there over the original Wesley Food Pantry.

```
#Total Plots
{plot(seq(1,72), data$WEFP, type="l",col = "blue", xlab = "Month and Year",ylab ="Total
 Individuals",main = "Total (Wesley and Parkland)",ylim = c(200,2500),lwd = 2)
  lines(seq(31,69), data$PFP[31:69], type="l",col = "red",lw=2)
  lines(seq(1,72), data$Total,type='l',col = "green",lw =2)
  legend("topright", c("Wesley","Parkland","Total"), lty = c(1, 1,1),lw=2,
col = c("blue", "red","green"), cex = .8)
  }
```



We see that even though the total number of people going to a food pantry is increasing, Wesley and Parkland are converging. This means that the general population in the CU area is electing to go to Parkland more and more often each year over Wesley. This inverse proportion may indicate that Parkland will soon overtake Wesley in number of people attending.

```
#Regression Model
{plot(seq(1,72), data$Total, type="l",col = "blue", xlab = "Month and Year",ylab ="Total
 Individuals",main = "Total Counts",ylim = c(600,2500),lwd=2)
  abline(lm(data$Total~seq(1,72)), col="red",lty=2,lw=1.5)}
```



Conclusion

Though our statistical tests were inconclusive, our graphs gave us a lot of insight into the data. Because we were able to visualize these trends, we can make some predictions about the future.

1. Parkland food pantry will likely overtake Wesley food pantry in size sometime in the next year
2. Months with Major holidays like Novemeber and December will continue to be busy each year
3. We can loosely predict the number of individuals that will come to the food pantry in a given month using the regression model.

```
#T-Tests
parkland  = data$PFP[31:69]
wesley  = data$WEFP[31:69]
t.test(parkland,wesley,paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  parkland and wesley
## t = -8.3921, df = 38, p-value = 3.511e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -498.1134 -304.5019
## sample estimates:
## mean of the differences
##               -401.3077
```

```
t.test(parkland,wesley)
```

```
##
##  Welch Two Sample t-test
##
## data:  parkland and wesley
## t = -9.0658, df = 74.972, p-value = 1.12e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -489.4910 -313.1244
## sample estimates:
## mean of x mean of y
##  534.8974  936.2051
```

```
#Regression Analysis
lm(formula = data$Total ~ seq(1, 72))
```

```
##
## Call:
## lm(formula = data$Total ~ seq(1, 72))
##
## Coefficients:
## (Intercept)   seq(1, 72)
##    1145.416        6.597
```

```
summary(lm(data$Total~seq(1,72)))
```

```
##
## Call:
## lm(formula = data$Total ~ seq(1, 72))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -462.57 -177.60   -7.72   99.60 1327.40
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1145.416     70.894  16.157  < 2e-16 ***
## seq(1, 72)     6.597      1.717   3.841 0.000294 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 282.1 on 61 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.1948, Adjusted R-squared:  0.1816
## F-statistic: 14.75 on 1 and 61 DF,  p-value: 0.000294
```