Final Project - Milestone 4

Dasun Wellawalage

November 11 2019

# Importing Data  
cost <- read.csv("C:/Users/Dasun/Desktop/Cost\_of\_Living\_Data.csv", stringsAsFactors = FALSE)  
  
# "Number.of.Workers" & "Age.of.Adults" seems like good candidates for factors. Converting them to factors  
cost['Number.of.Workers'] <- as.factor(cost$Number.of.Workers)  
cost['Age.of.Adults'] <- as.factor(cost$Age.of.Adults)  
  
# Removing Trailing spaces from the Area name  
cost$Area <- trimws(cost$Area, "r")  
  
# Loading required packages  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(broom)  
  
# Removing the "Period" & "Area" Columns as they add no value  
cost <- cost[c(3:16)]

1. Overall, write a coherent narrative that tells a story with the data as you complete this section.

The purpose if this project was to study living costs in across diffrent counties in Minnesota and figure out which factors affect the most when it comes to family expenses and to be able to make an informed decision about where to live based on the analysis. Data set is from 2016, so it might have changed somewhat as of today. Data set consisted of different variables such as the family size, how many workers are there in a family, how many children per family, total annual cost, and then a breakdown of those costs, such as housing, food, transport,health care, child care, tax etc. After cleaning up the data set, first I did a county wise analysis. Below are some of the interesting observations I had.

* Hourly wage seemed somewhat consistent across the state
* Taxes varied a lot across different counties. This is probably based on different neighborhoods i.e. rich neighborhoods will have more taxes
* CHild care expenses also varied greatly with the county. Some counties had almost double the average expenses
* Housing was mostly consistent while a few counties were significantly more expensinve than the rest

Housing seemed to have a significant impact on total expenses. Three counties turned up in top 5 for both most expensive counties and most expensive housing markets in Minnesota, and they were, Hennepin, Chisago, and Twin Cities. Surprisingly it was not the same when it comes to least expensinve counties. This shows that when it comes to basic living standards, housing doesn’t really plan an important role. Another important observation was that expensive counties had consistently higher hourly wages. Further, I did some predictive analysis as well. Current data set had an upper limit of $1400 for housing. As we know, it can even go up to $2000 in certain areas. When I performed a prediction on that it revealed that a person needs to have an annual income of approximately $120000 to afford a $2000 house in Minnesota.

Child care expenses also seemed to increase linearly with total expenses. However, in a vast majority of cases, child care expenses were less than $1000. An interesting observation was that although the number of children had an impact on family expenses, it didn’t necessarily increase the number of people working in a family. When the hourly wage was more than $20, families tend to have one working adult while those families with less than $15 had 2 working adults. I also made two interesting observations regarding the income and expenses of young adults vs those who are 51+. It seemed that in Minnesota, the average the minimum wage for young adults is $15 while the average maximum for 51+ adults is almost the same. However, this seems somewhat justified when I looked at the expenses where the average expenses for a 51+ person is less than a half of the average expenses for a young adult. Overall, I made some really interesting discoveries during my analysis.

1. Summarize the problem statement you addressed.

I was trying to address several issues related to lving in Minnesota here. Finding a suitable area to live, based on the living expenses and wages was the main problem I addressed. Apart from that I also addressed what the key concerns should be when calculating family expenses.

1. Summarize how you addressed this problem statement (the data used and the methodology employed).

# impact of Housing expenses on overall living cost  
lm(Yearly.Cost ~ Housing, data = cost)

##   
## Call:  
## lm(formula = Yearly.Cost ~ Housing, data = cost)  
##   
## Coefficients:  
## (Intercept) Housing   
## 2863.55 57.63

# Removing the top most housing data  
temp <- filter(cost, Housing < 1300)  
lm(Yearly.Cost ~ Housing, data = temp)

##   
## Call:  
## lm(formula = Yearly.Cost ~ Housing, data = temp)  
##   
## Coefficients:  
## (Intercept) Housing   
## -334.84 62.04

# Most expensive housing areas  
head(cost[order(-cost$Housing),])

## Family.Size Number.of.Adults Number.of.Children  
## 37 4 2 2  
## 38 4 2 2  
## 39 3 1 2  
## 40 4 2 2  
## 41 4 1 3  
## 42 5 2 3  
## Number.of.Workers Age.of.Adults Yearly.Cost Hourly.Wage  
## 37 2 19-50 90974 21.87  
## 38 1 full-time, 1 part-time 19-50 77406 24.81  
## 39 1 19-50 85239 40.98  
## 40 1 19-50 62011 29.81  
## 41 1 19-50 89623 43.09  
## 42 1 full-time, 1 part-time 19-50 81097 25.99  
## Child.Care Food Health.Care Housing Transport Other Taxes  
## 37 1680 989 417 1446 960 762 1327  
## 38 840 989 417 1446 960 762 1036  
## 39 1680 746 400 1446 897 686 1248  
## 40 0 989 417 1446 960 762 594  
## 41 1680 981 410 1446 960 759 1233  
## 42 840 1197 427 1446 1022 827 999

# Althouhg this data set has the higherst housing cost around 1400, it can even be around 2000 in some neighborhoods  
new <- data.frame(Housing = 2000)  
predict(lm(Yearly.Cost ~ Housing, data = cost), newdata = new)

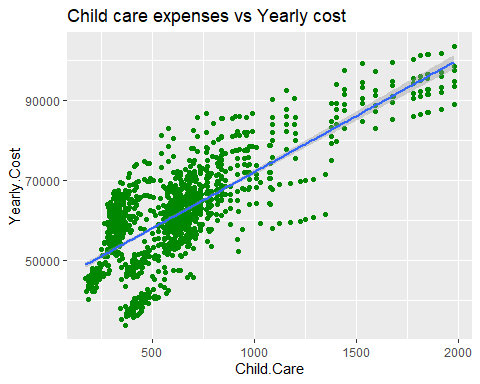
## 1   
## 118126.6

# How much do you have to make per hour to afford a $2000 house in Minnesota  
new <- data.frame(Yearly.Cost = 118126.6)  
predict(lm(Hourly.Wage ~ Yearly.Cost, data = cost), newdata = new)

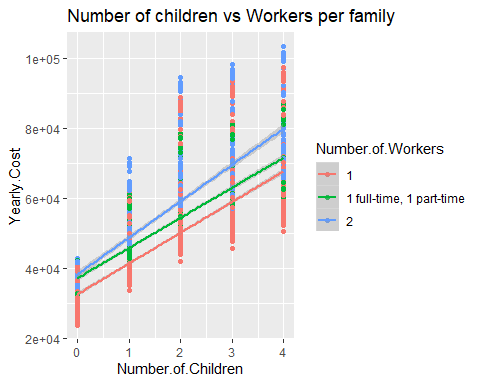
## 1   
## 39.08961

------------  
  
# impact of Child care expenses on overall living cost  
filter(cost, Child.Care > 0) %>% ggplot(aes(x = Child.Care, y = Yearly.Cost)) + geom\_point(position = "jitter", color = "#008800") + geom\_smooth(method = lm) + ggtitle("Child care expenses vs Yearly cost")

## Warning in Ops.factor(left): '-' not meaningful for factors  
  
## Warning in Ops.factor(left): '-' not meaningful for factors



# Does the number of children in a family have an impact on the number of people working  
ggplot(cost, aes(x = Number.of.Children, y = Yearly.Cost, color = Number.of.Workers)) + geom\_point() + geom\_smooth(method = lm) + ggtitle("Number of children vs Workers per family")



# Direct impact of hourly wage on Taxes  
lm(Taxes ~ Hourly.Wage, data = cost)

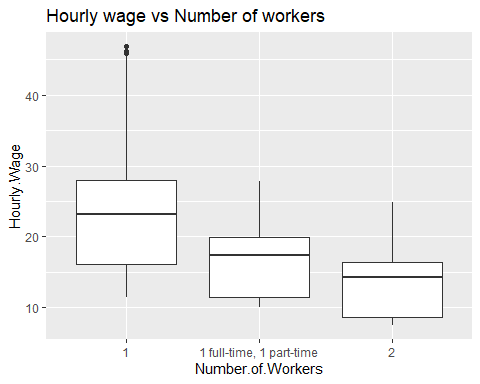
##   
## Call:  
## lm(formula = Taxes ~ Hourly.Wage, data = cost)  
##   
## Coefficients:  
## (Intercept) Hourly.Wage   
## 299.32 11.15

# How it changes as the nnumber of dependents (children) increase  
lm(Taxes ~ Hourly.Wage + Number.of.Children, data = cost)

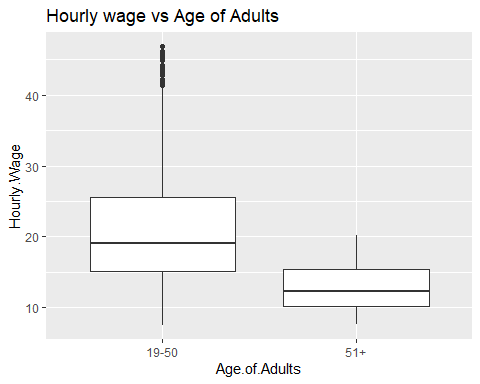
##   
## Call:  
## lm(formula = Taxes ~ Hourly.Wage + Number.of.Children, data = cost)  
##   
## Coefficients:  
## (Intercept) Hourly.Wage Number.of.Children   
## 347.143 4.468 47.059

1. Summarize the interesting insights that your analysis provided.

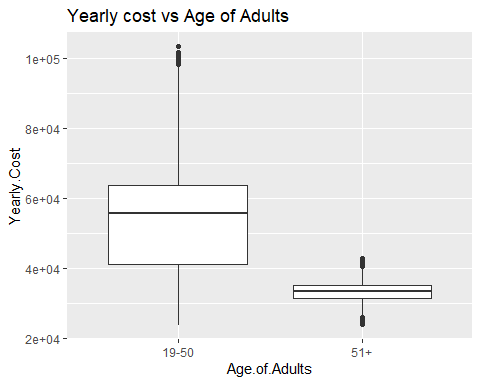
# relationship between the wages and number of workers per family  
ggplot(cost, aes(x =Number.of.Workers, y = Hourly.Wage)) + geom\_boxplot() + ggtitle("Hourly wage vs Number of workers")



# Those who are older than 51 years get paid significantly less than those who are younger  
ggplot(cost, aes(x =Age.of.Adults, y = Hourly.Wage)) + geom\_boxplot() + ggtitle("Hourly wage vs Age of Adults")



# How people in different ages manage their expenses  
ggplot(cost, aes(x =Age.of.Adults, y = Yearly.Cost)) + geom\_boxplot() + ggtitle("Yearly cost vs Age of Adults")



1. Summarize the implications to the consumer (target audience) of your analysis.

Depending on your age you can figure out Minnesota is the right state for you. It seems much better for younger people. housing costs and other costs seem pretty standard across the states. A few counties have high housing expenses though. Child care expenses and Taxes seem to vary a lot across the state. These 2 factors seem to be the most important when choosing an area to live in Minnesota.

1. Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.

* It is difficult to make decisions about the entire state based on this data set as we don’t have information for the entire country to compare with.
* Some more information related to the job category would have helped the audience make a much informed decision as different professionals have different salary scales.
* Another data set with savings information might give a better idea about how well people are doing in Minnesota. Wages alone won’t give you the complete picture about the financial status of a family. One can use the hourly wage calculate the annual salary and the take the difference between that and yearly cost to find out the annual savings, which can give more meaningful insight into the financial status of a family.

1. In addition, submit your completed Project using R Markdown or provide a link to where it can also be downloaded from and/or viewed.