MATH 130 Data Analysis Final Project

Shady Shamy

2/21/2021

Introduction

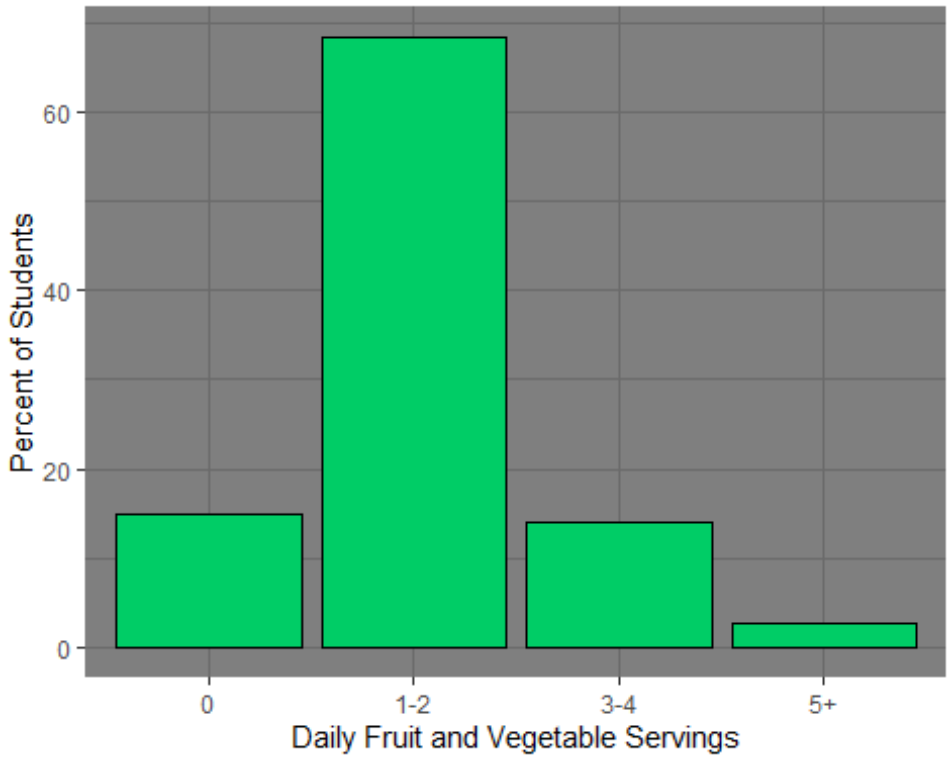
This data set was obtained by surveying 414 college students questions related to basic needs, including questions about CalFresh and eating patterns. The goal of this project is to identify if there is a relationship between monthly household income and fruit/vegetable consumption and to identify if college students receiving CalFresh benefits consume more fruits/vegetables compared to students not receiving CalFresh benefits.

```
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(readxl)
library(sjPlot)
library(RColorBrewer)
library(forcats)
chc <- read_excel("/Users/shady/Downloads/CHC Intern/data.xlsx", sheet = 1,
col_names = TRUE)
chc_fruit <- chc %>% select(fruit_veg_serv_daily) %>% na.omit
chc_monthly <- chc %>% select(monthly_hh_income) %>% na.omit

# Here, I omitted NA values separately for monthly household income and fruit
and vegetable consumption for univariate analysis.

chc_monthly_fruit_withNA <- select(chc, monthly_hh_income,
fruit_veg_serv_daily)
chc_monthly_fruit_withNA$chc_monthly_new <-
fct_collapse(chc_monthly_fruit_withNA$monthly_hh_income, "$2082 or Less" =
c("$150 or less", "$151-$2,082"), "$2083 - $3556" = c("$2,083-$2,820",
"$2,821-$3,556"), "$3557-$5030" = c("$3,557-$4,292", "$4,293-$5,030"),
"$5031-$6344" = c("$5,031-$5,694", "$5,695-$6,344"), "$6,345 or More" =
```

```
c("$6,345-$7,064", "$7,065 or more"))

chc_monthly_naomit <- chc_monthly_fruit_withNA %>% select(chc_monthly_new)
%>% na.omit
chc_fruit_monthly <- chc_monthly_fruit_withNA %>% select(chc_monthly_new,
fruit_veg_serv_daily) %>% na.omit

# Here, I selected monthly household income and fruit and vegetable servings
from the larger data set. I collapsed the monthly household income into 5
categories instead of 10 which makes it easier to read on a bar chart. After,
I made two new data sets. One with NA values omitted in monthly household
income for univariate analysis, and one with all NA values omitted for both
monthly household income and fruit and vegetable servings for bivariate
analysis.

chc$calfresh_new <- ifelse(chc$calfresh == 1, "Receives CalFresh", "Does Not
Receive CalFresh")
chc_calfresh_fruit <- chc %>% select(calfresh_new, fruit_veg_serv_daily) %>%
na.omit()
calfresh_naomit <- chc %>% select(calfresh_new) %>% na.omit()

# The CalFresh data was given as numbers with 1 being "yes" and 2 being "no"
so I created a factor variable with 1 being "Receives CalFresh" and 0 being
"Does Not Receive CalFresh." I then omitted NA values in all fruit and
vegetable consumption and Calfresh data for bivariate analysis. I also crated
a separate data set with just CalFresh NA data omitted for univariate
analysis.
```

Univariate Exploration:

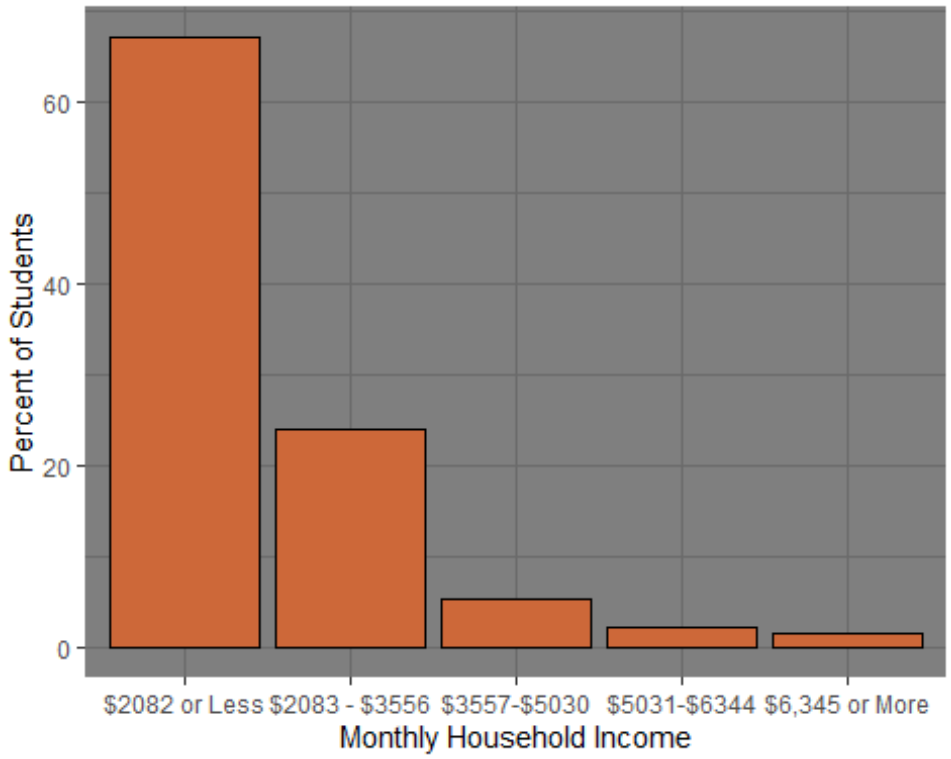
```
ggplot(chc_fruit, aes(x = fruit_veg_serv_daily)) + geom_bar(aes(y =
(..count..)*100/sum(..count..)), fill = "springgreen3", color = "black") +
theme_dark() + xlab("Daily Fruit and Vegetable Servings") + ylab("Percent of
Students")
```




```
table(chc$fruit_veg_serv_daily, useNA = "always")

##
##    0  1-2  3-4   5+ <NA>
##   51  234   48    9   72
```

The majority of students eat 1-2 servings of fruits and vegetables per day. This is under the recommended intake of 5 servings per day. 72 NA values were omitted.

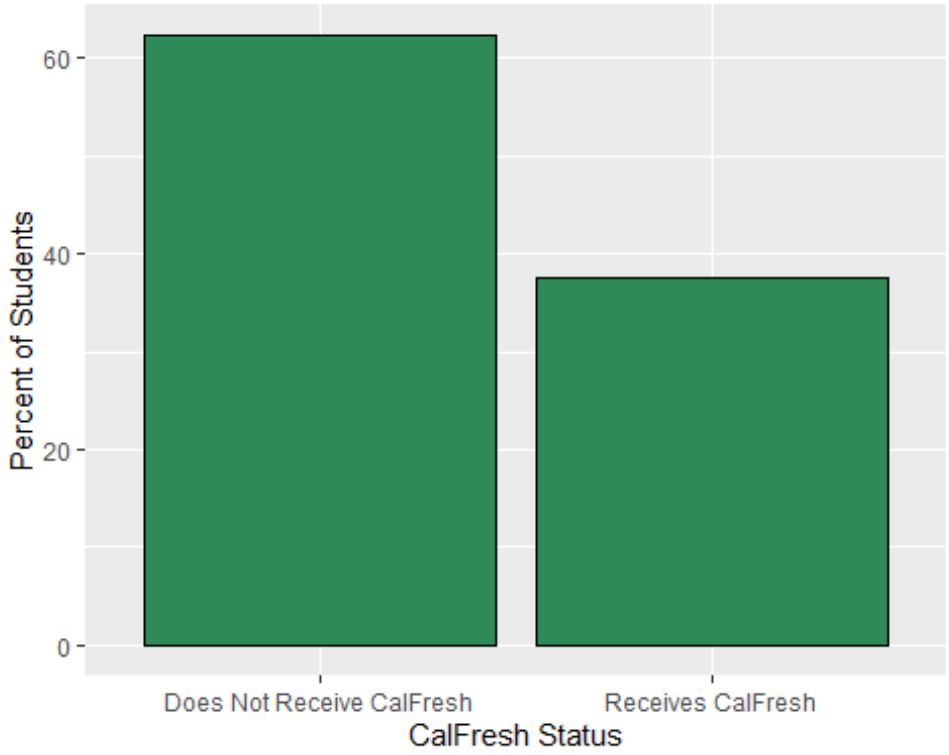
```
ggplot(chc_monthly_naomit, aes(x = chc_monthly_new)) + geom_bar(aes(y =
(..count..)*100/sum(..count..)), fill = "sienna3", color = "black") +
theme_dark() + ylab("Percent of Students") + xlab("Monthly Household Income")
```




```
table(chc_monthly_fruit_withNA$chc_monthly_new, useNA = "always")

##
##  $2082 or Less  $2083 - $3556    $3557-$5030    $5031-$6344 $6,345 or More
##            216             77             17              7             5
##           <NA>
##             92
```

The majority of students have a monthly household income of $2082 or less. 92 NA values were omitted.

```
ggplot(calfresh_naomit, aes(x = calfresh_new)) + geom_bar(aes(y =
(..count..)*100/sum(..count..)), fill="seagreen", color = "black") +
ylab("Percent of Students") + xlab("CalFresh Status")
```

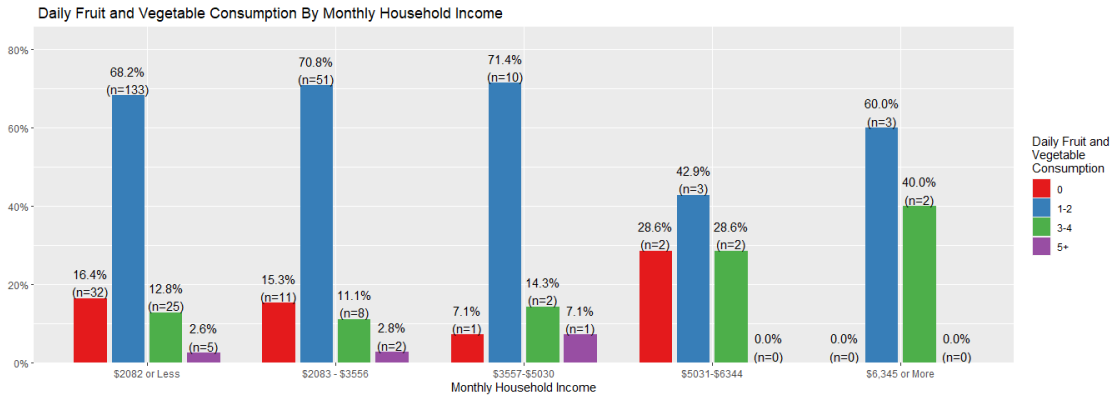



```
table(chc$calfresh_new, useNA = "always")

##
## Does Not Receive CalFresh         Receives CalFresh
<NA>
##                       237                       143
34
```

The majority of students surveyed did not receive any CalFresh benefits. 34 NA values were omitted.

Bivariate Exploration

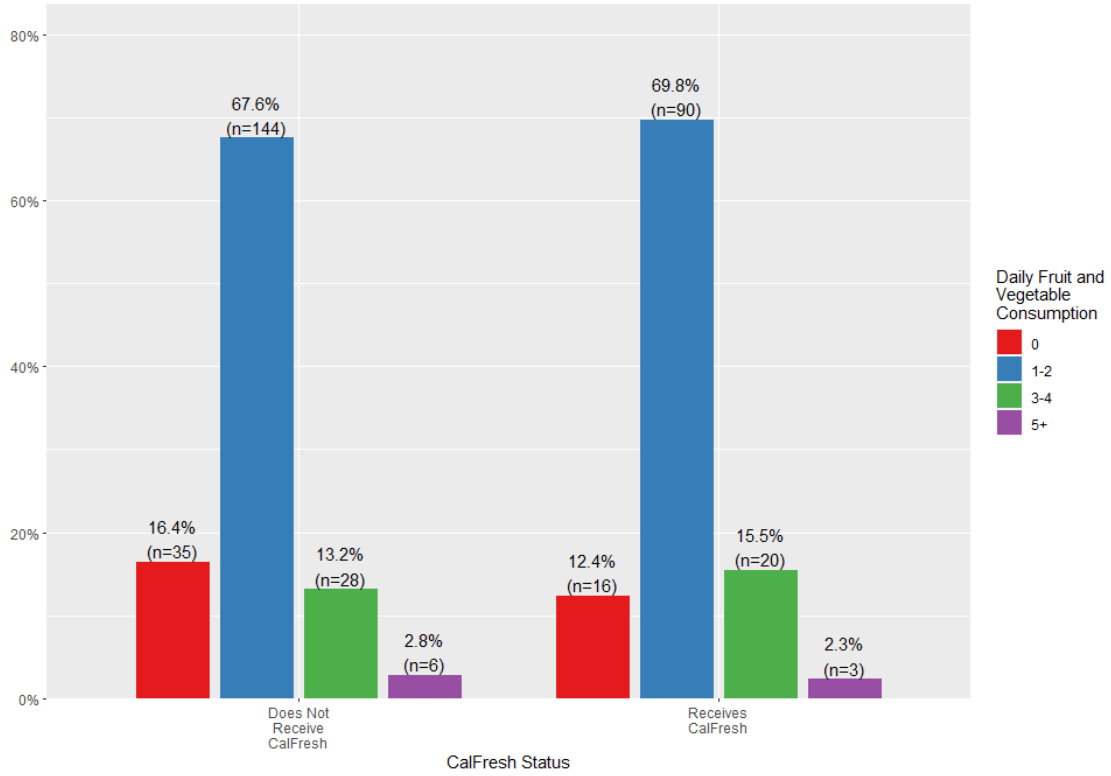
```
plot_xtab(chc_fruit_monthly$chc_monthly_new,
chc_fruit_monthly$fruit_veg_serv_daily, margin = 'row', title = " Daily Fruit
and Vegetable Consumption By Monthly Household Income", legend.title = "Daily
Fruit and Vegetable Consumption", axis.title = "Monthly Household Income",
wrap.title = 70, geom.colors = "Set1")
```




```
table(chc_monthly_fruit_withNA$chc_monthly_new,
chc_monthly_fruit_withNA$fruit_veg_serv_daily, useNA = "always")

##
##                     0 1-2 3-4  5+ <NA>
##   $2082 or Less    32 133  25   5   21
##   $2083 - $3556    11  51   8   2    5
##   $3557-$5030       1  10   2   1    3
##   $5031-$6344       2   3   2   0    0
##   $6,345 or More    0   3   2   0    0
##   <NA>              5  34   9   1   43
```

There seems to be no clear relationship between monthly household income and fruit and vegetable consumption for college students. 121 NA values were omitted.

```
plot_xtab(chc_calfresh_fruit$calfresh_new,
chc_calfresh_fruit$fruit_veg_serv_daily, margin = 'row', legend.title =
"Daily Fruit and Vegetable Consumption", axis.title = "CalFresh Status",
wrap.title = 70, geom.colors = "Set1")
```




```
table(chc$calfresh_new, chc$fruit_veg_serv_daily, useNA = "always")

##
##                                0 1-2 3-4  5+ <NA>
##   Does Not Receive CalFresh   35 144  28   6   24
##   Receives CalFresh           16  90  20   3   14
##   <NA>                         0   0   0   0   34
```

The amount 0f people eating 0 servings of fruits and vegetables per day decreased by 2.0% when receiving CalFresh benefits. The amount of people eating 3 or more servings increased by 1.8% when receiving CalFresh benefits. 72 NA values were omitted.

Conclusion

The majority of students have a monthly household income of $2082 or less, eat 1-2 servings of fruits and vegetables per day, and less than 40% of students receive CalFresh benefits. There is no clear relationship between monthly household income and fruit and vegetable consumption among college students. However, receiving CalFresh benefits may be increasing fruit and vegetable intake.