Study on

"Obesity Prevalence by Occupation in Washington State, Behavioral Risk Factor Surveillance System"

IS4250 HEALTHCARE ANALYTICS

Prepared by: Group 21

Guo Zhihan (A0099758B)

Zheng Yuan (A0105786X)

Table of Content

1 Introduction

Obesity has been declared as an epidemic, posing a threat to public health in the 21st century. It not only increases the number of people getting chronical diseases that affect the quality and the length of people's life, but also rises the potential medical costs of each individual and family, causing heavy financial burden to the whole society.

Based on the data provided by U.S department of health and human services, two thirds of adults are overweight or obese ("Healthy and Fit Nation Fact Sheet, " n.d.). In 2014, there were around 59.5% of U.S population employed in U.S, taking up more than half of the U.S population (Drew, 2014). On average, full-time workers spend more than 8 hours per day at work, with one-third to one-half of their time spent sitting down. In the 8-hour work, workers are likely to have one meal in the workplace. Hence, we find our interest in finding out the obesity prevalence study with a focus on the workforce population.

The paper, "Obesity Prevalence by Occupation in Washington State, Behavioral Risk Factor Surveillance System", is the first paper leveraging on the data collected by Behavioral Risk Factor Surveillance System to estimate the obesity prevalence by occupation in Washington State, USA. Based on the guideline from the writing studio of Duke University ("How to Review a Scientific Journal Article," n.d.), our report will firstly introduce the paper with a summary of the paper's purpose, assumptions, methodology as well as results and findings. We will also recognize the potential contributions can be made by the paper and address the existing challenges and limitations of the paper.

2 Summary

2.1 Purpose

The purpose of the paper is to estimate the prevalence of obesity by occupations and to examine the associations between obesity and demographic characteristics, a range of health behaviors as well as the occupational physical activity, therefore, to identify the occupations in which workers are at high risk of obesity in Washington State. By this study, government, policy makers are able to design and promote effective workplace wellness program especially for those occupations with high prevalence with obesity, with a direction of enhancing the identified health behavior factors that has significant association with obesity.

2.2 Questions

In this paper, there are a few questions being asked. The main questions to be answered is what the obesity prevalence of all categorized occupations are and what the occupations with significantly high obesity prevalence are. Furthermore, given a list of health behaviors, whether certain health behaviors have high association with prevalence of obesity was asked. Regarding to the method used in this paper, the measurement for obesity, the available data source and exception in the sample data was also asked. Another question is how to quantify the health behavior. Based on the result of the paper, whether there are any significant findings in the obesity prevalence by occupations, demographic characteristics and health behaviors was asked, as well as whether the result of the paper is by chance or consists with other research conducted.

2.3 Method

2.3.1 Data

Due to the characteristics of cross sectional study, it is very expensive to collect first-hand data for the study. In this paper, the authors made use of the data collected from the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS), which is an annual, state-based, random-digit-dialed, landline telephone survey on health conditions and behaviors of the noninstitutionalized US civilian population aged 18 years and older.

The reason of selecting the Washington State BRFSS survey data as the sample is the data includes the occupations of currently employed respondents since 2002. Since BRFSS also provides a biennial estimates of physical activity levels at work and during leisure time and assessment of health eating, the authors choose data from 2003 to 2009 every odd year. Furthermore, workers who were older than 65 years, who had a BMI of less than 18.5 were and who were from military or extraction occupations were excluded from the study.

As this survey also provides an annual national and state-level prevalence estimates of obesity and biennial estimates of physical activity levels at work and during leisure time and assessment of health eating, it provides sufficient variables for this paper.
In this paper, the authors used BMI to measure obesity, which can be calculated using respondents' self-reported weights and heights. Data about health behavior such as the intake of fruits and vegetables, smoke habit and time for leisure time physical activities were also

used. The authors then used the recommended levels of fruits and vegetables consumptions and leisure time physical activities from Healthy People 2020 to quantify the adequate fruits and vegetables intake and vigorous leisure time physical activities. Based on six questions asked related to fruits and vegetables intake, the authors calculated the average fruit and vegetables consumption per day. If it is larger than 5, it is categorized as having adequate fruit and vegetables. Having leisure-time physical activities 20 minutes or more a day for 3 or more days in a week is considered as having vigorous leisure-time physical activity.

In the sampling design, based on the Council of American Survey and Research Organization Guidelines, data were weighted according to the survey response rates and cooperation rate. The analysis performed in this paper is done by Stata software version 8.0 SE.

In this paper, the detailed calculation of prevalence ratio is not mentioned. Prevalence ratio is generally used to describe the percentage of a certain group of people in whole population. Therefore, with existing BRFSS data to calculate the prevalence ratio in our simulation, we use this following formula which is documented in Center of Disease Control and Prevention to simulate the calculation of obesity prevalence:

$$\text{Prevalence} = \frac{\text{persons with a given health indicator during a specified time period}}{\text{Population during the same time period}} \times 100$$

2.3.2 Analysis

In this paper, there is a discussion based on the result computed. The authors compared the result of this paper with a few similar researches. It is found that the occupations with significantly higher prevalence estimates of obesity are similar with those in other studies which analyzed the data collected from the National Health Interview Survey (NHIS). The authors also compared the result of the paper with an Australian National Health Survey on the finding of obesity prevalence associated with the amount of the occupational physical activity. In addition, the authors used the result from other studies to explain the possible causality for some of the findings. For example, the paper refers to the result of the research on socioeconomic status and obesity to explain the result of obesity prevalence with different socioeconomic status indicated in this research.

2.4 Assumption

There are some assumptions made in this paper. The first assumption made by the paper is BMI is an effective measure of obesity. There are many measures of obesity existing, some are simple calculations and some are very sophisticated. BMI is the most basic and most common method that are used in many research studies and this paper also used BMI as a measure of obesity. Although there are limitations of using BMI which will be covered in later part, BMI can be considered as an effective and cost-efficient measure of BMI.

Moreover, the paper also assumes that there is no biennial pattern on prevalence of obesity by occupation. Since BRFSS only provides a biennial estimates of physical activity levels at work and during leisure time and assessment of health eating, the paper has to target on 2003 to 2009 every odd year and assume that this set of odd year data is statistically reliable and representative.

Furthermore, the paper also assumes that similar occupations have similar characteristics and can be grouped together. From more than 80000 respondents, the paper identified 501 detailed occupational classes and further grouped them into 28 broader occupational categories. Authors of the paper assumed that this grouping is statistically.

2.5 Results and Findings

Out of 88121 respondents surveyed by Washington State BRFSS during the 4 years, 37626 respondents' survey data were valid because these respondents were aged between 18 to 64 years, had valid occupational code, had BMI equal or above 18.5 and were not from military or extraction occupations.

The overall prevalence of obesity for all occupations is 24.6% (95% CI, 24.0-25.1) ranging from 11.6% (95% CI, 8.0-15.2) for health diagnosing occupations to 38.6% (95% CI, 33.3-44.0) for truck drivers.

A significant disparity in prevalence of obesity of all occupations is observed. There are some occupations with significantly higher or lower prevalence of obesity. Three occupations with the highest prevalence of obesity are truck drivers (38.6%, 95% CI, 33.3-44.0), transportation and material moving (37.9%, 95% CI, 31.9-43.9) and protective services (33.3%, 95% CI, 28.5-38.1). Three occupations with the lowest prevalence of obesity are health diagnosing occupations (11.6%, 95% CI, 8.0-15.2), natural scientists and social scientists (17.3%, 95% CI, 12.9-21.7) and postsecondary teachers (17.6%, 95% CI, 12.4-22.7).

Prevalence of obesity is also related to demographic characters. The paper studied obesity prevalence in terms of some characteristics such as age, sex, annual household income and education level. Males have higher prevalence obesity (25.1%) than females (23.9%). Among three age groups, age between 45-64 years have highest prevalence of obesity of 27.2%, followed by age between 20-44 years (25.1%). Youngest workers have lowest prevalence of obesity of 18.3%. Among three annual household income groups, workers from high income household have lowest obesity prevalence of 22.3%. Workers from middle and low income household have roughly the same obesity prevalence of 26.9% and 26.2%. Obesity prevalence of workers with different education level also showed a similar pattern. Workers with college degree or higher have the lowest prevalence of obesity (20.5%). Workers with high school or some college education and less than Grade 12 education have obesity prevalence of 28.0% and 24.1% respectively.

One important finding from relationship between prevalence of obesity and demographic characteristics is workers with higher socioeconomic status tended to have lower prevalence of obesity than those with lower socioeconomic status. This finding can be explained that workers with high socioeconomic status are well educated, therefore, they care more about their physical appearance and health. Moreover, their higher income also allows them to buy expensive health products.

The report shows relationships between prevalence of obesity by occupation and occupational physical activity level and health behavioral factors. Truck drivers have highest proportion of smokers (34.1%, 95% CI, 28.6-39.4) and health assessment and treating, excluding registered nurses have the lowest proportion of smokers (3.2%, 95% CI, 0.9-5.7). Health diagnosing occupations (38.2%, 95% CI, 32.8-43.7) and health assessment and treating, excluding registered nurses (37.5%, 95% CI, 31.5-43.4) have highest proportion of workers who have adequate fruit and vegetable intake. Mechanics and repairers, mathematical and computer scientists, and truck drivers have the lowest proportion of people having adequate fruit and vegetable intake. The highest proportion of vigorous LTPA was among workers in protective services (50.8%; 95% CI, 45.7–56.0), followed by health diagnosing occupations (45.3%; 95% CI, 39.6–50.9) and postsecondary teachers (42.7%; 95%

CI, 35.8–49.6); machine operators, assemblers, and inspectors had the lowest proportion (27.2%; 95% CI, 22.4–31.9). More than 80% of workers in cleaning and building services, construction and construction trades, and farming, forestry, and fishing reported their work as physically demanding, but 3% or less of lawyers and judges and teachers, excluding postsecondary felt that their occupation is physically demanding.

Since prevalence of obesity was associated with demographic characteristics, health behavioral factors and occupational physical activity level, the report shows the prevalence ratios of different categories in these factors. Prevalence ratios (PRs) for obesity were significantly higher among workers in older age groups than among workers aged 18 to 29, among male workers than among female workers, and among workers with less education than among workers with a college degree or higher. Compared with workers in the highest income group (≥$75,000), those in the lowest income group (<$35,000) had significantly higher obesity prevalence (PR = 1.09, 95% CI, 1.03–1.16), but the difference in PR was not significant for workers in the middle income group ($35, 000–$74,999). Nonsmokers were more likely to be obese (PR = 1.17, 95% CI, 1.09–1.25) than smokers. Workers who had adequate daily consumption of fruits and vegetables and adequate LTPA had significantly lower prevalence of obesity compared with those who consumed and exercised less (PR = 0.91, 95% CI, 0.86–0.97 and PR = 0.63, 95% CI, 0.60–0.67, respectively). Workers who characterized their occupational physical activity level as physically demanding had a lower prevalence of obesity (PR = 0.83; 95% CI, 0.78–0.88) than those who characterized their occupational physical activity level as non-physically demanding.

One important finding from the report is that workers who smoke had lower prevalence of obesity than those who do not smoke. This phenomenon can be explained by the appetite-suppressing effects of smoking. However, this finding seems contradict with the report that truck drivers have highest proportion of workers who are smokers but truck drivers still have highest prevalence of obesity. This can be explained that occupation is the aggregate measure of all the demographic, socioeconomic, health behavioral factor. Although smoking has a protective effect on obesity, this effect does not offset the effect of other factors on obesity. For example, truck drivers usually have irregular mealtime and limited availability of food, leading to frequent consumption of fast food for truck drivers. Therefore, truck drivers still have highest prevalence of obesity even though they also have highest proportion of smokers.

Apart from smoking, the report also shows that adequate fruit and vegetable consumption and LTPA level have protective effect on obesity.

3 Contribution

Firstly, this paper computes the obesity prevalence ratio for the workforce in Washington State by occupations, demographic characteristics, health behaviors and physical activities. There is no other paper discussing about the prevalence by occupation when the paper was published. While workforce in U.S weighs more than half of the whole population, this study provides valuable insight on the obesity issue in this big group.

Secondly, the paper may contribute to increasing awareness in obesity prevalence in workplace. This paper was supported by the Washington State Department of Labor and Industries and NIOSH. The result of the paper indicates the current situation of workforce obesity existing in Washington State, and hopefully will increase the awareness on obesity issue of the relevant department of Washington state governments and policy makers.

Thirdly, effective interventions from governments, companies and policy makers can be carried out based on the societal and occupational factors associated with obesity identified in the paper. The result of the paper is able to show key factors that significantly affect the obesity. The evidence provides additional information for people working on the design and promotion of workplace health program in the society. Policy makers are able to identify the industries or groups that require more public health resources allocated based on the finding of the paper. As obesity leads to the more costs associated with sick leave and absenteeism in the workplace as well as the increasing risks of developing chronic diseases, effective implementation of these workplace health programs will reduce the associated costs for the company as well as the health care costs of individuals and health care system in a long run.

Last but not least, this paper provides an example for other researchers to study on the obesity prevalence by occupation in a national level or by other geographical locations, so that specified factors may be identified in the implementation of workplace health program catering for the particular group of population.

4 Limitations

The paper is subject to four major limitations. In term of the nature of the study, there are two main limitations. Firstly, since it is a cross-sectional study, the results only show correlation between obesity and occupation but does not provide relationship of causation. Secondly, using BMI as a measure of obesity may be biased because BMI measures cannot differentiate fat and lean tissue mass; employees from physically demanding jobs are physically fit but they have higher BMIs because of higher proportion of muscle in their bodies. For example, employees from protective services have the highest proportion of vigorous LTPA (50.8%, 95% CI, 45.7-56.0) but also have the third highest prevalence of obesity (33.3%, 95% CI, 28.5-38.1). There are also two major limitations in term of the data used. Firstly, BRFSS uses respondents' self-reported measures of height and weight rather than actual measure of height and weight, which may result in an underestimation of respondents' BMIs. Furthermore, since BRFSS only collected data from landline telephone households and English- or Spanish-speaking respondents, there might be systematic errors because cellular phone-only households were not included in the survey.

Although these limitations mentioned exist, BRFSS is still be most optimal system to collect data on state-level health-related behaviors by occupations because of the large sample size and cost-effectiveness of the data collection process.

5 Simulation

We simulated part of the study using the data downloaded from Centers for Disease Control and Prevention in USA. Our simulation focus on calculating the obesity prevalence ratio by demographic characteristics (sex, race, annual house income, education attachment) and health behaviors and physical activities (smoke, level of fruit and vegetables intake, level of leisure-time physical activities, occupational physical activities type). This study only took a small part of the BRFSS data while the data originally has 405 variables covering responses from all the states in U.S. Due to the huge size of raw data file, we decided to process only year 2009 of the data out of total 4 years.

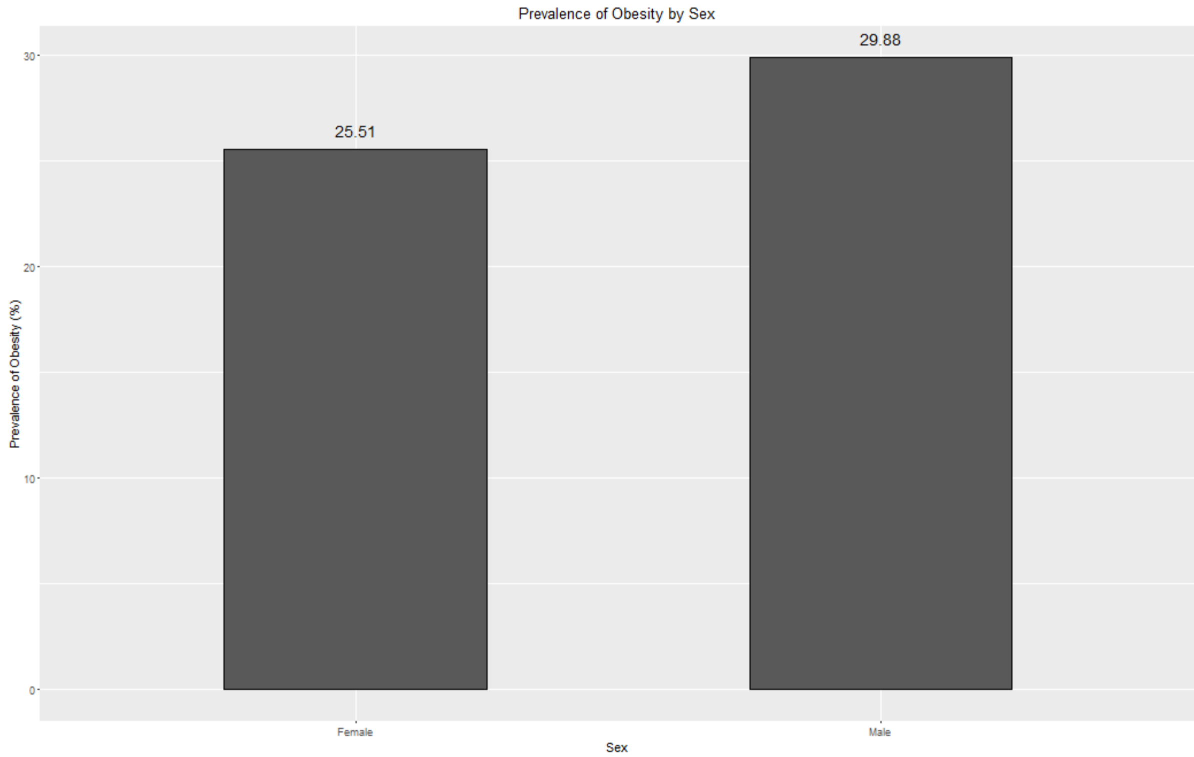
The data file for 2009 contains 432607 rows and we filtered them out based on the selection requirement in our studying paper. Below is the code sample of how we extracted the sample data.

```r
## import data
origin_data_09 <- read.xport("CDBRFS09.XPT")
View(origin_data_09)

sample_data_09 <- dplyr::select(origin_data_09, X_STATE, VETERAN2,  AGE, EMPLOY, HTM3, WTKG2
    , SEX, RACE2, INCOME2, EDUCA, SMOKDAY2, JOBACTIV, VIGCAT_, X_FRTINDX, X_BMI4)
sample_data_09 <- dplyr::rename(sample_data_09, VETERAN = VETERAN2)
sample_data_09 <- dplyr::filter(sample_data_09, X_STATE == 53, EMPLOY == 1 | EMPLOY == 2)


sample_data_all <- dplyr::filter(sample_data_09, VETERAN != 1 & VETERAN != 2 & VETERAN != 3)
sample_data_all <- dplyr::filter(sample_data_all, AGE >= 18 & AGE <= 64, X_BMI4 >= 1850 &
    X_BMI4 != 9999 )
```

Taking prevalence ratio by sex as an example, we filtered out responses indicating sex as female in the sample and saved as *female_data* and then filtered obese people among *female_data*. We used count in dplyr package to count the number of obese female and total female and used the prevalence ratio formula to do the calculation. Last but not least, we used dataframe to present data in a table format and applied ggplot package to visualize the table.

```r
## Prevelance Ratio by Sex

female_data <- dplyr::filter(sample_data_all, SEX==2)
female_data_obese <- dplyr::filter(female_data, X_BMI4 >= 3000)

female_obese <- dplyr::count(female_data_obese)
female_total <- dplyr::count(female_data)
female_pr = female_obese / female_total

male_data <- dplyr::filter(sample_data_all, SEX==1)
male_data_obese <- dplyr::filter(male_data, X_BMI4 >= 3000)

male_obese <- dplyr::count(male_data_obese)
male_total <- dplyr::count(male_data)
male_pr = male_obese / male_total

ob_pr_sex_female = female_pr/female_pr
ob_pr_sex_male_female = male_pr/female_pr

ob_pr_sex <- dplyr::bind_rows(female_pr , male_pr)
ob_pr_sex_adjusted <- dplyr::bind_rows(ob_pr_sex_female, ob_pr_sex_male_female)

sex = c("Female","Male")
result_sex=data.frame(sex, ob_pr_sex, ob_pr_sex_adjusted)
names(result_sex)[2] <- "Prevalence Ratio"
names(result_sex)[3] <- "Adjusted Prevalence Ratio"

     sex Prevalence Ratio Adjusted Prevalence Ratio
1 Female        0.2703583                  1.000000
2   Male        0.2869479                  1.061361
```

```r
## Prevelance Ratio by Sex
result_sex
colnames(result_sex) <- c("Prevalence Ratio", "pr")
result_sex$pr=as.numeric(format(round(result_sex$pr,4), nsmall=4))
ggplot(result_sex, aes(sex,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Sex")+ylab("
    Prevalence of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of
    Obesity by Sex")
```



There are limitations in our simulation. Firstly, we used the sample data that was indicated in the reference of the report, however, we did not find any information about occupation of respondents. We did email Washington state for the state BRFSS data, however, we still not able to find the data on occupation. Hence, we are not able to perform the obesity prevalence by occupation. Secondly, due to the missing data above, we are not able to remove respondents working in extraction and military. To make our sample data as close as possible to the data in the paper, we used the question on military training in order to exclude people working in military.

In the simulation, most of the trends are similar to what was presented in the paper. Most of results computed by us will be included as part of the appendix.

6 Conclusion

Our report has reviewed the purpose, research question and assumptions of the study and introduced our simulation for the study. We have also reviewed the data source and methodologies used in their study. Then we discussed the results and some important findings of the study. In the last section, we have also acknowledged the contribution of the paper and addressed some limitations and challenges of their study.

7 Reference

This report is based on the study conducted by David K. Bonauto, Dayu Lu and Joyce Fan: http://www.cdc.gov/pcd/issues/2014/13_0219.htm

1. The Surgeon General's Vision for a Healthy and Fit Nation Fact Sheet. (n.d.). Retrieved from http://www.surgeongeneral.gov/priorities/healthy-fit-nation/obesityvision_factsheet.html

2. Drew, D. (2014, November 7). Employment, unemployment and underemployment: Different stories from the jobs numbers. Retrieved from http://www.pewresearch.org/fact-tank/2014/11/07/employment-vs-unemployment-different-stories-from-the-jobs-numbers/

3. How to Read and Review a Scientific Journal Article: Writing Summaries and Critiques. (n.d.). Retrieved from: http://twp.duke.edu/uploads/media_items/scientificarticlereview.original.pdf

8 Appendix

A. Simulation output

a. Overall obesity prevalence

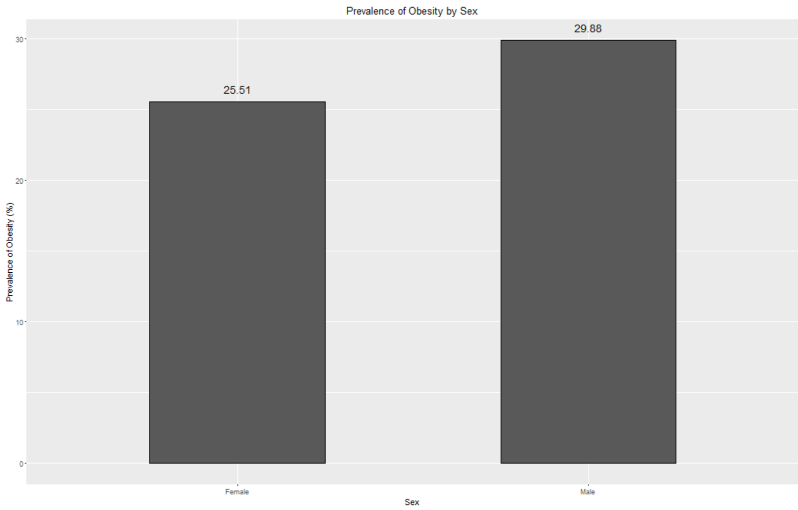
```
## Sample data Prevalence Ratio
result_all

##   all_occupations Prevalence Ratio
## 1             All        0.2771925
```

b. Obesity prevalence by sex

```
## Prevelance Ratio by Sex
result_sex

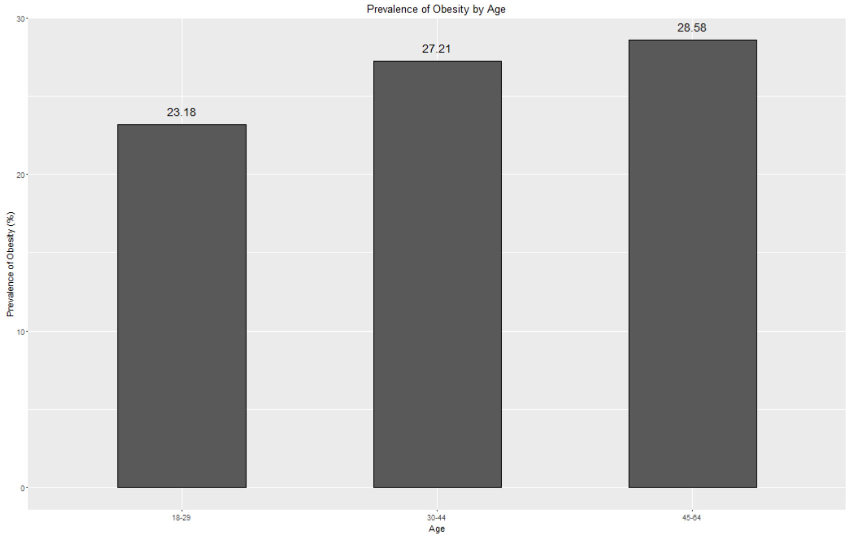
##      sex Prevalence Ratio Adjusted Prevalence Ratio
## 1 Female        0.2703583                  1.000000
## 2   Male        0.2869479                  1.061361
```



c. Obesity prevalence by age

```
## Prevelance Ratio by Age
result_age

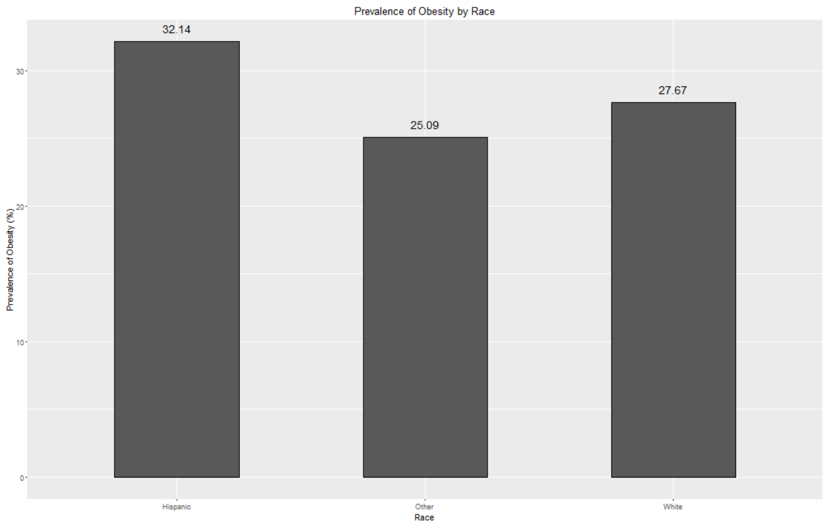
##     age Prevalence Ratio Adjusted Prevalence Ratio
## 1 18-29        0.2317881                  1.000000
## 2 30-44        0.2720655                  1.173768
## 3 45-64        0.2857777                  1.232927
```



d. Obesity prevalence by race

```
## Prevelance Ratio by Race
result_race

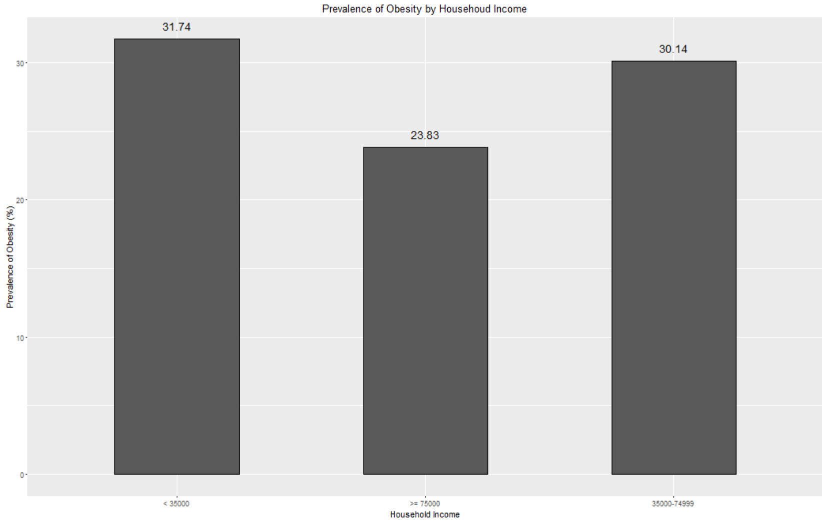
##       race Prevalence Ratio Adjusted Prevalence Ratio
## 1    White        0.2766767                 1.0000000
## 2 Hispanic        0.3214286                 1.1617479
## 3    Other        0.2508591                 0.9066868
```



e. Obesity prevalence by annual household income

```
## Prevelance Ratio by Annual Household Income
result_income

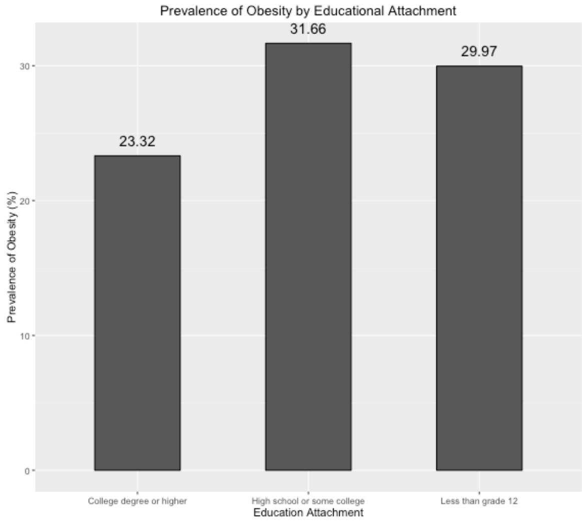
##   household_income Prevalence Ratio Adjusted Prevalence Ratio
## 1          < 35000        0.3174377                  1.332012
## 2      35000-74999        0.3035646                  1.273798
## 3         >= 75000        0.2383144                  1.000000
```



f. Obesity prevalence by education attachment

```
## Prevelance Ratio by Education Attachment
result_education

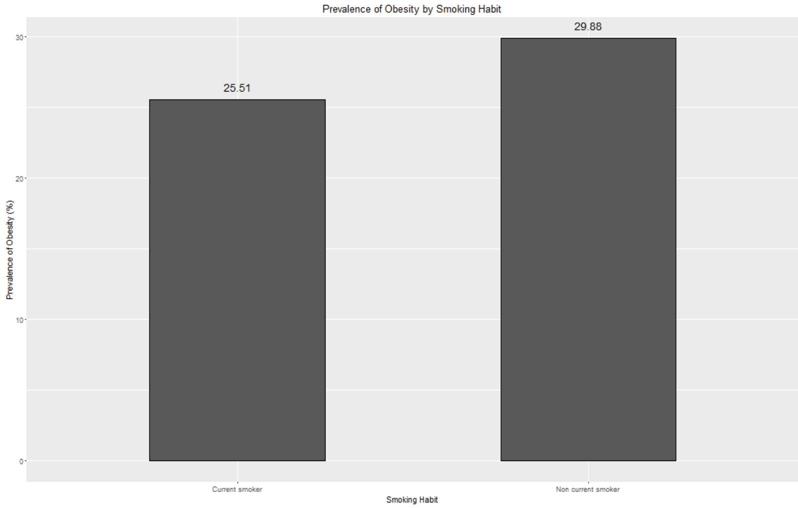
##          education_attachment         n      n.1
## 1          Less than grade 12 0.2996633 1.285138
## 2 High school or some college 0.3165746 1.357664
## 3    College degree or higher 0.2331759 1.000000
```



g. Obesity prevalence by smoke

```
## Prevelance Ratio by Smoke
result_smoke

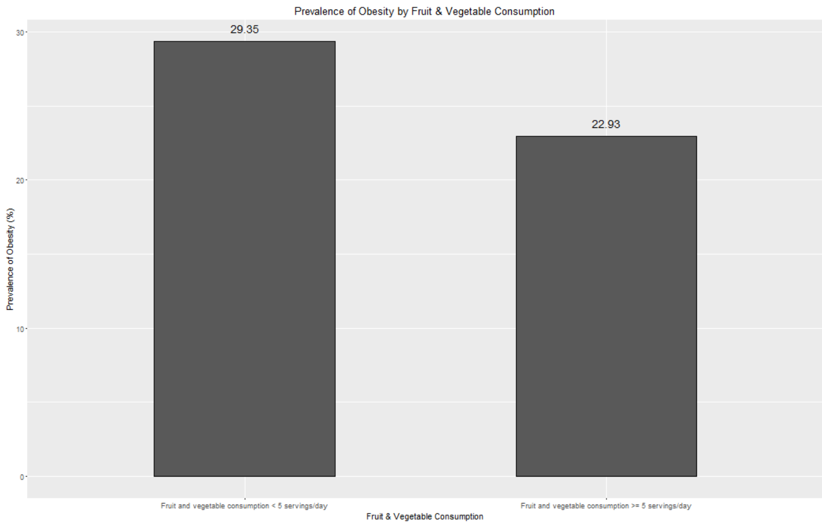
##        smoking_habit Prevalence Ratio Adjusted Prevalence Ratio
## 1     Current smoker        0.2551230                  1.000000
## 2 Non current smoker        0.2988134                  1.171252
```



h. Obesity prevalence by fruits and vegetables intakes

```
## Prevelance Ratio by level of Fruits and Vegetables intakes
result_fruit

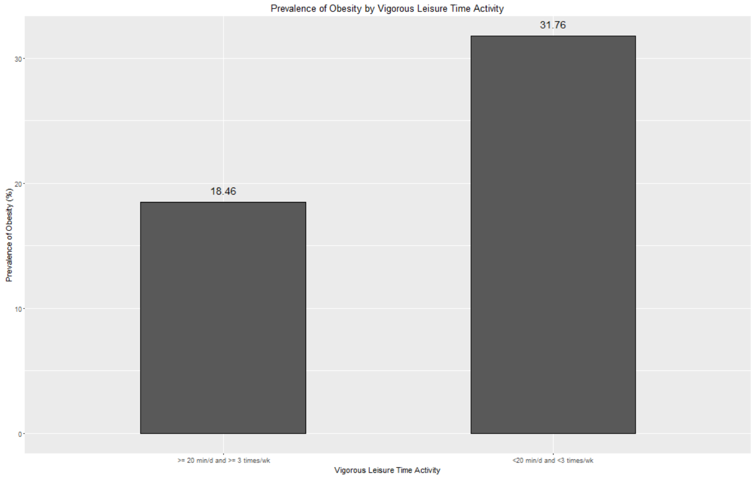
##                       fruits_vegetables_consumption Prevalence Ratio
## 1  Fruit and vegetable consumption < 5 servings/day        0.2934722
## 2 Fruit and vegetable consumption >= 5 servings/day        0.2292892
##   Adjusted Prevalence Ratio
## 1                 1.0000000
## 2                 0.7812977
```



i. Obesity prevalence by leisure-time physical activities

```
## Prevelance Ratio by level of Leisure-time Physical Activities
result_LTPA

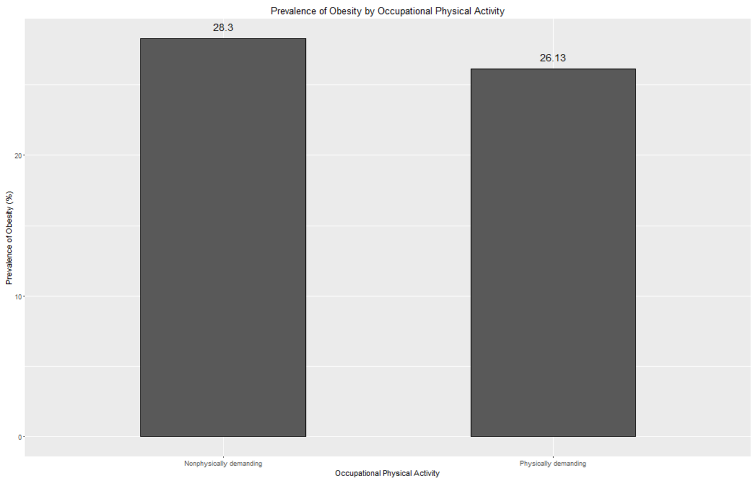
##   VigorousLeisureTimePhysicalActivity Prevalence Ratio
## 1           <20 min/d and <3 times/wk        0.3175772
## 2       >= 20 min/d and >= 3 times/wk        0.1845878
##   Adjusted Prevalence Ratio
## 1                 1.0000000
## 2                 0.5812375
```



j. Obesity prevalence by occupational physical activities

```
## Prevelance Ratio by Occupational Physical Activities
result_OPA

##   occupationalPhysicalActivity Prevalence Ratio Adjusted Prevalence Ratio
## 1      Nonphysically demanding        0.2829642                 1.0000000
## 2         Physically demanding        0.2613438                 0.9235932
```



B. Source code

```
---
title: "IS4250/Group 21"
output: md_document
---

```{r}
##Please SCROLL down to see the output tablend charts

library(utils)
library(dplyr)
library(foreign)
library(ggplot2)

## import data
origin_data_09 <- read.xport("CDBRFS09.XPT")

sample_data_09 <- dplyr::select(origin_data_09, X_STATE, VETERAN2,  AGE, EMPLOY, HTM3, WTKG2,
SEX, RACE2, INCOME2, EDUCA, SMOKDAY2, JOBACTIV, VIGCAT_, X_FRTINDX, X_BMI4)
sample_data_09 <- dplyr::rename(sample_data_09, VETERAN = VETERAN2)
sample_data_09 <- dplyr::filter(sample_data_09, X_STATE == 53, EMPLOY == 1 | EMPLOY == 2)


sample_data_all <- dplyr::filter(sample_data_09, VETERAN != 1 & VETERAN != 2 & VETERAN != 3)
sample_data_all <- dplyr::filter(sample_data_all, AGE >= 18 & AGE <= 64, X_BMI4 >= 1850 &
X_BMI4 != 9999 )

## Sample data
all_obese <- dplyr::filter(sample_data_all, X_BMI4 >= 3000)
all_obese_count <- dplyr::count(all_obese)
all_count <- dplyr::count(sample_data_all)
all_pr = all_obese_count / all_count

all_occupations = c("All")
result_all = data.frame(all_occupations, all_pr)
names(result_all)[2] <- "Prevalence Ratio"

```

```{r}
## Prevelance Ratio by Sex

female_data <- dplyr::filter(sample_data_all, SEX==2)
female_data_obese <- dplyr::filter(female_data, X_BMI4 >= 3000)

female_obese <- dplyr::count(female_data_obese)
female_total <- dplyr::count(female_data)
female_pr = female_obese / female_total

male_data <- dplyr::filter(sample_data_all, SEX==1)
male_data_obese <- dplyr::filter(male_data, X_BMI4 >= 3000)

male_obese <- dplyr::count(male_data_obese)
male_total <- dplyr::count(male_data)
male_pr = male_obese / male_total

ob_pr_sex_female = female_pr/female_pr
ob_pr_sex_male_female = male_pr/female_pr

ob_pr_sex <- dplyr::bind_rows(female_pr , male_pr)
ob_pr_sex_adjusted <- dplyr::bind_rows(ob_pr_sex_female, ob_pr_sex_male_female)

sex = c("Female","Male")
result_sex=data.frame(sex, ob_pr_sex, ob_pr_sex_adjusted)
names(result_sex)[2] <- "Prevalence Ratio"
names(result_sex)[3] <- "Adjusted Prevalence Ratio"
```
```

````
```

```{r}
## Prevelance Ratio by Age
young_age_data <- dplyr::filter(sample_data_all, AGE >= 18 & AGE <= 29)
young_age_data_obese <- dplyr::filter(young_age_data, X_BMI4 >= 3000)
young_obese <- dplyr::count(young_age_data_obese)
young_total <- dplyr::count(young_age_data)
young_age_pr = young_obese/young_total

middle_age_data <- dplyr::filter(sample_data_all, AGE >= 30 & AGE <= 44)
middle_age_data_obese <- dplyr::filter(middle_age_data, X_BMI4 >= 3000)
middle_obese <- dplyr::count(middle_age_data_obese)
middle_total <- dplyr::count(middle_age_data)
middle_age_pr = middle_obese/middle_total

old_age_data <- dplyr::filter(sample_data_all, AGE >= 45 & AGE <= 64)
old_age_data_obese <- dplyr::filter(old_age_data, X_BMI4 >= 3000)
old_obese <- dplyr::count(old_age_data_obese)
old_total <- dplyr::count(old_age_data)
old_age_pr = old_obese/old_total

ob_pr_age_young = young_age_pr/young_age_pr
ob_pr_age_middle_young = middle_age_pr/young_age_pr
ob_pr_age_old_young = old_age_pr/young_age_pr

ob_pr_age <- dplyr::bind_rows(young_age_pr, middle_age_pr)
ob_pr_age <- dplyr::bind_rows(ob_pr_age,old_age_pr)

ob_pr_age_adjusted <- dplyr::bind_rows(ob_pr_age_young,ob_pr_age_middle_young)
ob_pr_age_adjusted <- dplyr::bind_rows(ob_pr_age_adjusted,ob_pr_age_old_young)

age = c("18-29", "30-44", "45-64")
result_age = data.frame(age, ob_pr_age, ob_pr_age_adjusted)
names(result_age)[2] <- "Prevalence Ratio"
names(result_age)[3] <- "Adjusted Prevalence Ratio"

```

```{r}
## Prevelance Ratio by Race

white_data <- dplyr::filter(sample_data_all, RACE2 == 1)
white_data_obese <- dplyr::filter(white_data, X_BMI4 >= 3000)
white_obese <- dplyr::count(white_data_obese)
white_total <- dplyr::count(white_data)
white_pr = white_obese/white_total

hispanic_data <- dplyr::filter(sample_data_all, RACE2 == 8)
hispanic_data_obese <- dplyr::filter(hispanic_data, X_BMI4 >= 3000)
hispanic_obese <- dplyr::count(hispanic_data_obese)
hispanic_total <- dplyr::count(hispanic_data)
hispanic_pr = hispanic_obese/hispanic_total

other_race_data <- dplyr::filter(sample_data_all, RACE2 != 8 & RACE2 != 1)
other_race_data_obese <- dplyr::filter(other_race_data, X_BMI4 >= 3000)
other_race_obese <- dplyr::count(other_race_data_obese)
other_race_total <- dplyr::count(other_race_data)
other_race_pr = other_race_obese/other_race_total

ob_pr_age_white = white_pr/white_pr
ob_pr_age_hispanic = hispanic_pr/white_pr
ob_pr_age_other_race = other_race_pr/white_pr

ob_pr_race <- dplyr::bind_rows(white_pr, hispanic_pr)
ob_pr_race <- dplyr::bind_rows(ob_pr_race, other_race_pr)
````

````
ob_pr_race_adjusted <- dplyr::bind_rows(ob_pr_age_white,ob_pr_age_hispanic)
ob_pr_race_adjusted <- dplyr::bind_rows(ob_pr_race_adjusted,ob_pr_age_other_race)

race = c("White", "Hispanic", "Other")
result_race = data.frame(race, ob_pr_race, ob_pr_race_adjusted)
names(result_race)[2] <- "Prevalence Ratio"
names(result_race)[3] <- "Adjusted Prevalence Ratio"

```


```{r}
## Prevelance Ratio by Annual Household Income

income <- dplyr::filter(sample_data_all, INCOME2 != 77 & INCOME2 != 99)

low_income_data <- dplyr::filter(income, INCOME2 < 6)
low_income_data_obese <- dplyr::filter(low_income_data, X_BMI4 >= 3000)
low_income_obese <- dplyr::count(low_income_data_obese)
low_income_total <- dplyr::count(low_income_data)
low_income_pr = low_income_obese/low_income_total

middle_income_data <- dplyr::filter(income, INCOME2 >= 6 & INCOME2 < 8)
middle_income_data_obese <- dplyr::filter(middle_income_data, X_BMI4 >= 3000)
middle_income_obese <- dplyr::count(middle_income_data_obese)
middle_income_total <- dplyr::count(middle_income_data)
middle_income_pr = middle_income_obese/middle_income_total

high_income_data <- dplyr::filter(income, INCOME2 >= 8)
high_income_data_obese <- dplyr::filter(high_income_data, X_BMI4 >= 3000)
high_income_obese <- dplyr::count(high_income_data_obese)
high_income_total <- dplyr::count(high_income_data)
high_income_pr = high_income_obese/high_income_total

ob_pr_income_high = high_income_pr/high_income_pr
ob_pr_income_middle_high = middle_income_pr/high_income_pr
ob_pr_income_low_high = low_income_pr/high_income_pr

ob_pr_income <- dplyr::bind_rows(low_income_pr, middle_income_pr)
ob_pr_income <- dplyr::bind_rows(ob_pr_income,high_income_pr)


ob_pr_income_adjusted <- dplyr::bind_rows(ob_pr_income_low_high, ob_pr_income_middle_high)
ob_pr_income_adjusted <- dplyr::bind_rows(ob_pr_income_adjusted,ob_pr_income_high)

household_income = c("< 35000", "35000-74999", ">= 75000")
result_income = data.frame(household_income, ob_pr_income, ob_pr_income_adjusted)
names(result_income)[2] <- "Prevalence Ratio"
names(result_income)[3] <- "Adjusted Prevalence Ratio"

```

```{r}
## Prevelance Ratio by Education Attachment

education <- dplyr::filter(sample_data_all, EDUCA != 9)

low_edu_data <- dplyr::filter(education, EDUCA < 4)
low_edu_data_obese <- dplyr::filter(low_edu_data, X_BMI4 >= 3000)
low_edu_obese <- dplyr::count(low_edu_data_obese)
low_edu_total <- dplyr::count(low_edu_data)
low_edu_pr = low_edu_obese/low_edu_total

middle_edu_data <- dplyr::filter(education, EDUCA >= 4 & EDUCA <6)
middle_edu_data_obese <- dplyr::filter(middle_edu_data, X_BMI4 >= 3000)
middle_edu_obese <- dplyr::count(middle_edu_data_obese)
````

````
middle_edu_total <- dplyr::count(middle_edu_data)
middle_edu_pr = middle_edu_obese/middle_edu_total

high_edu_data <- dplyr::filter(education, EDUCA == 6)
high_edu_data_obese <- dplyr::filter(high_edu_data, X_BMI4 >= 3000)
high_edu_obese <- dplyr::count(high_edu_data_obese)
high_edu_total <- dplyr::count(high_edu_data)
high_edu_pr = high_edu_obese/high_edu_total

ob_pr_edu_high = high_edu_pr/high_edu_pr
ob_pr_edu_middle_high = middle_edu_pr/high_edu_pr
ob_pr_edu_low_high = low_edu_pr/high_edu_pr

ob_pr_edu <- dplyr::bind_rows(low_edu_pr, middle_edu_pr)
ob_pr_edu <- dplyr::bind_rows(ob_pr_edu,high_edu_pr)
ob_pr_edu_adjusted <- dplyr::bind_rows(ob_pr_edu_low_high, ob_pr_edu_middle_high)
ob_pr_edu_adjusted <- dplyr::bind_rows(ob_pr_edu_adjusted,ob_pr_edu_high)

education_attachment = c("Less than grade 12", "High school or some college", "College degree
or higher")
result_education = data.frame(education_attachment, ob_pr_edu, ob_pr_edu_adjusted)

```

```{r}
## Prevelance Ratio by Smoke

smoke <- dplyr::filter(sample_data_all, SMOKDAY2 == 1 | SMOKDAY2 == 2)
smoke_obese_data <- dplyr::filter(smoke, X_BMI4 >= 3000)

smoke_obese <- dplyr::count(smoke_obese_data)
smoke_total <- dplyr::count(smoke)
smoke_pr = smoke_obese/smoke_total

nonsmoke <- dplyr::filter(sample_data_all, SMOKDAY2 != 9 & SMOKDAY2 != 1 & SMOKDAY2 != 2)
nonsmoke_obese_data <- dplyr::filter(nonsmoke, X_BMI4 >= 3000)

nonsmoke_obese <- dplyr::count(nonsmoke_obese_data)
nonsmoke_total <- dplyr::count(nonsmoke)
nonsmoke_pr = nonsmoke_obese / nonsmoke_total


ob_pr_smoke = smoke_pr/smoke_pr
ob_pr_nonsmoke_smoke = nonsmoke_pr/smoke_pr

ob_pr_smoke_table <- dplyr::bind_rows(smoke_pr, nonsmoke_pr)
ob_pr_smoke_table_adjusted <- dplyr::bind_rows(ob_pr_smoke , ob_pr_nonsmoke_smoke)

smoking_habit = c("Current smoker", "Non current smoker")
result_smoke = data.frame(smoking_habit, ob_pr_smoke_table, ob_pr_smoke_table_adjusted)
names(result_smoke)[2] <- "Prevalence Ratio"
names(result_smoke)[3] <- "Adjusted Prevalence Ratio"

```

```{r}
## Prevelance Ratio by level of Fruits and Vegetables intakes

frequentFruit <- dplyr::filter(sample_data_all, X_FRTINDX == 4)
frequentFruit_obese_data <- dplyr::filter(frequentFruit, X_BMI4 >= 3000)

frequentFruit_obese <- dplyr::count(frequentFruit_obese_data)
frequentFruit_total <- dplyr::count(frequentFruit)
frequentFruit_pr = frequentFruit_obese/frequentFruit_total

nonfrequentFruit <- dplyr::filter(sample_data_all, X_FRTINDX == 1 | X_FRTINDX ==2 | X_FRTINDX
== 3)
````

```
nonfrequentFruit_obese_data <- dplyr::filter(nonfrequentFruit, X_BMI4 >= 3000)

nonfrequentFruit_obese <- dplyr::count(nonfrequentFruit_obese_data)
nonfrequentFruit_total <- dplyr::count(nonfrequentFruit)
nonfrequentFruit_pr = nonfrequentFruit_obese / nonfrequentFruit_total


ob_pr_nonfrequentFruit_nonfrequentFruit = nonfrequentFruit_pr/nonfrequentFruit_pr
ob_pr_frequentFruit_nonfrequentFruit = frequentFruit_pr/nonfrequentFruit_pr

ob_pr_frequentFruit_table <- dplyr::bind_rows(nonfrequentFruit_pr, frequentFruit_pr)
ob_pr_frequentFruit_table_adjusted <- dplyr::bind_rows(ob_pr_nonfrequentFruit_nonfrequentFruit,
ob_pr_frequentFruit_nonfrequentFruit)

fruits_vegetables_consumption = c("Fruit and vegetable consumption < 5 servings/day", "Fruit
and vegetable consumption >= 5 servings/day")
result_fruit = data.frame(fruits_vegetables_consumption, ob_pr_frequentFruit_table,
ob_pr_frequentFruit_table_adjusted)

names(result_fruit)[2] <- "Prevalence Ratio"
names(result_fruit)[3] <- "Adjusted Prevalence Ratio"
```

```{r}
## Prevelance Ratio by level of Leisure-time Physical Activities

vigorousLTPA <- dplyr::filter(sample_data_all, VIGCAT_ == 1)
vigorousLTPA_obese_data <- dplyr::filter(vigorousLTPA, X_BMI4 >= 3000)

vigorousLTPA_obese <- dplyr::count(vigorousLTPA_obese_data)
vigorousLTPA_total <- dplyr::count(vigorousLTPA)
vigorousLTPA_pr = vigorousLTPA_obese/vigorousLTPA_total

nonvigorousLTPA <- dplyr::filter(sample_data_all, VIGCAT_ == 2 | VIGCAT_ == 3)
nonvigorousLTPA_obese_data <- dplyr::filter(nonvigorousLTPA, X_BMI4 >= 3000)

nonvigorousLTPA_obese <- dplyr::count(nonvigorousLTPA_obese_data)
nonvigorousLTPA_total <- dplyr::count(nonvigorousLTPA)
nonvigorousLTPA_pr = nonvigorousLTPA_obese / nonvigorousLTPA_total


ob_pr_nonvigorousLTPA_nonvigorousLTPA = nonvigorousLTPA_pr/nonvigorousLTPA_pr
ob_pr_vigorousLTPA_nonvigorousLTPA = vigorousLTPA_pr/nonvigorousLTPA_pr

ob_pr_vigorousLTPA_table <- dplyr::bind_rows(nonvigorousLTPA_pr, vigorousLTPA_pr)
ob_pr_vigorousLTPA_table_adjusted <- dplyr::bind_rows(ob_pr_nonvigorousLTPA_nonvigorousLTPA,
ob_pr_vigorousLTPA_nonvigorousLTPA)

VigorousLeisureTimePhysicalActivity = c("<20 min/d and <3 times/wk", " >= 20 min/d and >= 3
times/wk")
result_LTPA = data.frame(VigorousLeisureTimePhysicalActivity, ob_pr_vigorousLTPA_table,
ob_pr_vigorousLTPA_table_adjusted)

names(result_LTPA)[2] <- "Prevalence Ratio"
names(result_LTPA)[3] <- "Adjusted Prevalence Ratio"
```

```{r}
## Prevelance Ratio by Occupational Physical Activities

physicalDemandJob <- dplyr::filter(sample_data_all, JOBACTIV == 2 | JOBACTIV == 3)
physicalDemandJob_obese_data <- dplyr::filter(physicalDemandJob, X_BMI4 >= 3000)

physicalDemandJob_obese <- dplyr::count(physicalDemandJob_obese_data)
physicalDemandJob_total <- dplyr::count(physicalDemandJob)
physicalDemandJob_pr = physicalDemandJob_obese/physicalDemandJob_total
```

```
nonphysicalDemandJob <- dplyr::filter(sample_data_all, JOBACTIV == 1)
nonphysicalDemandJob_obese_data <- dplyr::filter(nonphysicalDemandJob, X_BMI4 >= 3000)

nonphysicalDemandJob_obese <- dplyr::count(nonphysicalDemandJob_obese_data)
nonphysicalDemandJob_total <- dplyr::count(nonphysicalDemandJob)
nonphysicalDemandJob_pr = nonphysicalDemandJob_obese / nonphysicalDemandJob_total


ob_pr_nonphysicalDemandJob_nonphysicalDemandJob =
nonphysicalDemandJob_pr/nonphysicalDemandJob_pr
ob_pr_physicalDemandJob_nonphysicalDemandJob = physicalDemandJob_pr/nonphysicalDemandJob_pr

ob_pr_physicalDemandJob_table <- dplyr::bind_rows(nonphysicalDemandJob_pr,
physicalDemandJob_pr)
ob_pr_physicalDemandJob_table_adjusted <-
dplyr::bind_rows(ob_pr_nonphysicalDemandJob_nonphysicalDemandJob,
ob_pr_physicalDemandJob_nonphysicalDemandJob)

occupationalPhysicalActivity = c("Nonphysically demanding", "Physically demanding")
result_OPA = data.frame(occupationalPhysicalActivity, ob_pr_physicalDemandJob_table,
ob_pr_physicalDemandJob_table_adjusted)

names(result_OPA)[2] <- "Prevalence Ratio"
names(result_OPA)[3] <- "Adjusted Prevalence Ratio"
```

```{r}
## In below tables, the 2nd column represents the prevelance ratio of all the sample that
meets the requirement corresponding in 1st column. The 3 column is the adjusted ratio. If one
has a value 1 in that column, it means it is the reference ratio.

## Sample data Prevalence Ratio
result_all
## Prevelance Ratio by Sex
result_sex
colnames(result_sex) <- c("Prevalence Ratio", "pr")
result_sex$pr=as.numeric(format(round(result_sex$pr,4), nsmall=4))
ggplot(result_sex,
aes(sex,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Sex")+ylab("Prevalenc
e of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of
Obesity by Sex")
## Prevelance Ratio by Age
result_age
colnames(result_age) <- c("Prevalence Ratio", "pr")
result_age$pr=as.numeric(format(round(result_age$pr,4), nsmall=4))
ggplot(result_age,
aes(age,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Age")+ylab("Prevalenc
e of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of
Obesity by Age")
## Prevelance Ratio by Race
result_race
colnames(result_race) <- c("Prevalence Ratio", "pr")
result_race$pr=as.numeric(format(round(result_race$pr,4), nsmall=4))
ggplot(result_race,
aes(race,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Race")+ylab("Prevale
nce of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of
Obesity by Race")
## Prevelance Ratio by Annual Household Income
result_income
colnames(result_income) <- c("Prevalence Ratio", "pr")
result_income$pr=as.numeric(format(round(result_income$pr,4), nsmall=4))
ggplot(result_income,
aes(household_income,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Househol
d Income")+ylab("Prevalence of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1, size =
5)+ggtitle("Prevalence of Obesity by Househoud Income")
## Prevelance Ratio by Education Attachment
result_education
```

```
colnames(result_education) <- c("Prevalence Ratio", "pr")
result_education$pr=as.numeric(format(round(result_education$pr,4), nsmall=4))
ggplot(result_education,
aes(education_attachment,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Educ
ation Attachment")+ylab("Prevalence of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1,
size = 5)+ggtitle("Prevalence of Obesity by Educational Attachment")
## Prevelance Ratio by Smoke
result_smoke
colnames(result_smoke) <- c("Prevalence Ratio", "pr")
result_smoke$pr=as.numeric(format(round(result_smoke$pr,4), nsmall=4))
ggplot(result_smoke,
aes(smoking_habit,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xlab("Smoking
Habit")+ylab("Prevalence of Obesity (%)")+geom_text(aes(label = pr*100), vjust=-1, size =
5)+ggtitle("Prevalence of Obesity by Smoking Habit")
## Prevelance Ratio by level of Fruits and Vegetables intakes
result_fruit
colnames(result_fruit) <- c("Prevalence Ratio", "pr")
result_fruit$pr=as.numeric(format(round(result_fruit$pr,4), nsmall=4))
ggplot(result_fruit,
aes(fruits_vegetables_consumption,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+x
lab("Fruit & Vegetable Consumption")+ylab("Prevalence of Obesity (%)")+geom_text(aes(label =
pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of Obesity by Fruit & Vegetable Consumption")
## Prevelance Ratio by level of Leisure-time Physical Activities
result_LTPA
colnames(result_LTPA) <- c("Prevalence Ratio", "pr")
result_LTPA$pr=as.numeric(format(round(result_LTPA$pr,4), nsmall=4))
ggplot(result_LTPA,
aes(VigorousLeisureTimePhysicalActivity,pr*100))+geom_bar(colour="black",width=.5,stat="identi
ty")+xlab("Vigorous Leisure Time Activity")+ylab("Prevalence of Obesity
(%)")+geom_text(aes(label = pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of Obesity by
Vigorous Leisure Time Activity")
## Prevelance Ratio by Occupational Physical Activities
result_OPA
colnames(result_OPA) <- c("Prevalence Ratio", "pr")
result_OPA$pr=as.numeric(format(round(result_OPA$pr,4), nsmall=4))
ggplot(result_OPA,
aes(occupationalPhysicalActivity,pr*100))+geom_bar(colour="black",width=.5,stat="identity")+xl
ab("Occupational Physical Activity")+ylab("Prevalence of Obesity (%)")+geom_text(aes(label =
pr*100), vjust=-1, size = 5)+ggtitle("Prevalence of Obesity by Occupational Physical Activity")
```