Assignment 2 ANA 515

Daniel Appadurai

2023-06-11

# Heavy Drinking Among Adults In the US Aged 18 and Above

#importing dependencies  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)  
library(knitr)  
library(rmarkdown)  
library(officer)

#importing csv file and creating a dataframe named data. The dataframe is created with the read.csv() function from the base R package. Columns 1 to 6 of 34 and the top 5 rows of 'data' dataset are displayed.  
data<-read.csv("U.S.\_Chronic\_Disease\_Indicators\_\_CDI\_.csv")  
subset\_data1<-data[1:5,1:6]  
kable(subset\_data1)

| YearStart | YearEnd | LocationAbbr | LocationDesc | DataSource | Topic |
| --- | --- | --- | --- | --- | --- |
| 2014 | 2014 | AR | Arkansas | SEDD; SID | Asthma |
| 2018 | 2018 | CO | Colorado | SEDD; SID | Asthma |
| 2018 | 2018 | DC | District of Columbia | SEDD; SID | Asthma |
| 2017 | 2017 | GA | Georgia | SEDD; SID | Asthma |
| 2010 | 2010 | MI | Michigan | SEDD; SID | Asthma |

#Columns 7 to 12 of 34 and the top 5 rows of 'data' dataset are displayed.  
subset\_data2<-data[1:5,7:12]  
kable(subset\_data2)

| Question | Response | DataValueUnit | DataValueType | DataValue | DataValueAlt |
| --- | --- | --- | --- | --- | --- |
| Hospitalizations for asthma | NA |  | Number | 916 | 916 |
| Hospitalizations for asthma | NA |  | Number | 2227 | 2227 |
| Hospitalizations for asthma | NA |  | Number | 708 | 708 |
| Hospitalizations for asthma | NA |  | Number | 3520 | 3520 |
| Hospitalizations for asthma | NA |  | Number | 123 | 123 |

#Columns 13 to 16 of 34 and the top 5 rows of 'data' dataset are displayed.  
subset\_data3<-data[1:5,13:16]  
kable(subset\_data3)

| DataValueFootnoteSymbol | DatavalueFootnote | LowConfidenceLimit | HighConfidenceLimit |
| --- | --- | --- | --- |
|  |  | NA | NA |
|  |  | NA | NA |
|  |  | NA | NA |
|  |  | NA | NA |
|  |  | NA | NA |

#Columns 17 to 20 of 34 and the top 5 rows of 'data' dataset are displayed.  
subset\_data4<-data[1:5,17:20]  
kable(subset\_data4)

| StratificationCategory1 | Stratification1 | StratificationCategory2 | Stratification2 |
| --- | --- | --- | --- |
| Gender | Male | NA | NA |
| Overall | Overall | NA | NA |
| Overall | Overall | NA | NA |
| Gender | Female | NA | NA |
| Race/Ethnicity | Hispanic | NA | NA |

#Columns 21 to 24 of 34 and the top 5 rows of 'data' dataset are displayed.  
subset\_data5<-data[1:5,21:24]  
kable(subset\_data5)

| StratificationCategory3 | Stratification3 | GeoLocation | ResponseID |
| --- | --- | --- | --- |
| NA | NA | POINT (-92.27449074299966 34.74865012400045) | NA |
| NA | NA | POINT (-106.13361092099967 38.843840757000464) | NA |
| NA | NA | POINT (-77.036871 38.907192) | NA |
| NA | NA | POINT (-83.62758034599966 32.83968109300048) | NA |
| NA | NA | POINT (-84.71439026999968 44.6613195430005) | NA |

#Columns 25 to 30 of 34 and the top 5 rows of 'data' dataset are displayed.  
subset\_data6<-data[1:5,25:30]  
kable(subset\_data6)

| LocationID | TopicID | QuestionID | DataValueTypeID | StratificationCategoryID1 | StratificationID1 |
| --- | --- | --- | --- | --- | --- |
| 5 | AST | AST3\_1 | NMBR | GENDER | GENM |
| 8 | AST | AST3\_1 | NMBR | OVERALL | OVR |
| 11 | AST | AST3\_1 | NMBR | OVERALL | OVR |
| 13 | AST | AST3\_1 | NMBR | GENDER | GENF |
| 26 | AST | AST3\_1 | NMBR | RACE | HIS |

#Columns 31 to 34 of 34 and the top 5 rows of 'data' dataset are displayed.  
subset\_data7<-data[1:5,31:34]  
kable(subset\_data7)

| StratificationCategoryID2 | StratificationID2 | StratificationCategoryID3 | StratificationID3 |
| --- | --- | --- | --- |
| NA | NA | NA | NA |
| NA | NA | NA | NA |
| NA | NA | NA | NA |
| NA | NA | NA | NA |
| NA | NA | NA | NA |

# SECTION 1 AND 2:

## Chronic Disease Indicators Dataset:

The chronic disease indicators data set is a comprehensive data set collected on an annual basis since 2001 by the CDC’s Division of Population Health across all 50 US states. The data set is in the .csv file format and read into a dataframe using the read.csv() function from the base R package. The dataset has a total of 34 columns and 1185676 rows as of 2021 and contains 124 indicators for a total of 17 chronic diseases.

## Heavy Alcohol Consumption Among Adults Aged 18 And Above:

One of the chronic disease data collected is on heavy alcohol consumption among adults across 50 US states. This data analysis is done on heavy alcohol consumption for the year 2021. The analysis looks into heavy alcohol consumption among men and women and also among the various racial/ethnic groups.

# SECTION 3: DATA CLEANING

#Inspecting columns  
colnames(data)

## [1] "YearStart" "YearEnd"   
## [3] "LocationAbbr" "LocationDesc"   
## [5] "DataSource" "Topic"   
## [7] "Question" "Response"   
## [9] "DataValueUnit" "DataValueType"   
## [11] "DataValue" "DataValueAlt"   
## [13] "DataValueFootnoteSymbol" "DatavalueFootnote"   
## [15] "LowConfidenceLimit" "HighConfidenceLimit"   
## [17] "StratificationCategory1" "Stratification1"   
## [19] "StratificationCategory2" "Stratification2"   
## [21] "StratificationCategory3" "Stratification3"   
## [23] "GeoLocation" "ResponseID"   
## [25] "LocationID" "TopicID"   
## [27] "QuestionID" "DataValueTypeID"   
## [29] "StratificationCategoryID1" "StratificationID1"   
## [31] "StratificationCategoryID2" "StratificationID2"   
## [33] "StratificationCategoryID3" "StratificationID3"

# filtering data for the most recent year, 2021, same number of columns as the previous dataframe, but trucated to the first six for organized and clean output but the appropriate application of the code will be displayed in the output  
recent\_data<-filter(data,YearEnd==2021)  
subset\_recent\_data<-recent\_data[1:5, 1:6]  
kable(subset\_recent\_data)

| YearStart | YearEnd | LocationAbbr | LocationDesc | DataSource | Topic |
| --- | --- | --- | --- | --- | --- |
| 2021 | 2021 | ND | North Dakota | BRFSS | Arthritis |
| 2021 | 2021 | US | United States | BRFSS | Arthritis |
| 2021 | 2021 | NJ | New Jersey | BRFSS | Arthritis |
| 2021 | 2021 | OH | Ohio | BRFSS | Arthritis |
| 2021 | 2021 | VA | Virginia | BRFSS | Arthritis |
|  |  |  |  |  |  |

#inspecting the chronic disease topics  
unique(recent\_data$Topic)

## [1] "Arthritis"   
## [2] "Alcohol"   
## [3] "Tobacco"   
## [4] "Oral Health"   
## [5] "Asthma"   
## [6] "Diabetes"   
## [7] "Overarching Conditions"   
## [8] "Nutrition, Physical Activity, and Weight Status"  
## [9] "Cardiovascular Disease"   
## [10] "Mental Health"   
## [11] "Chronic Obstructive Pulmonary Disease"   
## [12] "Chronic Kidney Disease"   
## [13] "Immunization"   
## [14] "Reproductive Health"

# Filtering only for chronic disease category, alcohol,same number of columns as the previous dataframe, but trucated to the first six for organized and clean output but the appropriate application of the code will be displayed in the output  
alcohol<-filter(recent\_data,Topic=="Alcohol")  
subset\_alcohol<-alcohol[1:5,1:6]  
kable(subset\_alcohol)

| YearStart | YearEnd | LocationAbbr | LocationDesc | DataSource | Topic |
| --- | --- | --- | --- | --- | --- |
| 2021 | 2021 | AZ | Arizona | APIS | Alcohol |
| 2021 | 2021 | ME | Maine | Legal Research | Alcohol |
| 2021 | 2021 | VA | Virginia | APIS | Alcohol |
| 2021 | 2021 | MD | Maryland | APIS | Alcohol |
| 2021 | 2021 | WA | Washington | APIS | Alcohol |

#inspecting null columns and compling a list of all columns. The result of null columns will be TRUE and non-null columns will be FALSE  
results<- list()  
for (col\_name in colnames(alcohol)) {  
 analysis\_results<- all(is.na(alcohol[[col\_name]]))  
 results[[col\_name]]<- analysis\_results  
}

# Making a dataframe from the column list and transforming the shape of the dataframe for better view  
df<-data.frame(results)  
df2<-t(df)  
df3<-as.data.frame(df2)

## V1  
## YearStart FALSE  
## YearEnd FALSE  
## LocationAbbr FALSE  
## LocationDesc FALSE  
## DataSource FALSE  
## Topic FALSE  
## Question FALSE  
## Response TRUE  
## DataValueUnit FALSE  
## DataValueType FALSE  
## DataValue FALSE  
## DataValueAlt FALSE  
## DataValueFootnoteSymbol FALSE  
## DatavalueFootnote FALSE  
## LowConfidenceLimit FALSE  
## HighConfidenceLimit FALSE  
## StratificationCategory1 FALSE  
## Stratification1 FALSE  
## StratificationCategory2 TRUE  
## Stratification2 TRUE  
## StratificationCategory3 TRUE  
## Stratification3 TRUE  
## GeoLocation FALSE  
## ResponseID TRUE  
## LocationID FALSE  
## TopicID FALSE  
## QuestionID FALSE  
## DataValueTypeID FALSE  
## StratificationCategoryID1 FALSE  
## StratificationID1 FALSE  
## StratificationCategoryID2 TRUE  
## StratificationID2 TRUE  
## StratificationCategoryID3 TRUE  
## StratificationID3 TRUE

#Naming the row index as column name  
library(dplyr)  
df3<-rownames\_to\_column(df3, var="Column\_Name")  
df3

## Column\_Name V1  
## 1 YearStart FALSE  
## 2 YearEnd FALSE  
## 3 LocationAbbr FALSE  
## 4 LocationDesc FALSE  
## 5 DataSource FALSE  
## 6 Topic FALSE  
## 7 Question FALSE  
## 8 Response TRUE  
## 9 DataValueUnit FALSE  
## 10 DataValueType FALSE  
## 11 DataValue FALSE  
## 12 DataValueAlt FALSE  
## 13 DataValueFootnoteSymbol FALSE  
## 14 DatavalueFootnote FALSE  
## 15 LowConfidenceLimit FALSE  
## 16 HighConfidenceLimit FALSE  
## 17 StratificationCategory1 FALSE  
## 18 Stratification1 FALSE  
## 19 StratificationCategory2 TRUE  
## 20 Stratification2 TRUE  
## 21 StratificationCategory3 TRUE  
## 22 Stratification3 TRUE  
## 23 GeoLocation FALSE  
## 24 ResponseID TRUE  
## 25 LocationID FALSE  
## 26 TopicID FALSE  
## 27 QuestionID FALSE  
## 28 DataValueTypeID FALSE  
## 29 StratificationCategoryID1 FALSE  
## 30 StratificationID1 FALSE  
## 31 StratificationCategoryID2 TRUE  
## 32 StratificationID2 TRUE  
## 33 StratificationCategoryID3 TRUE  
## 34 StratificationID3 TRUE

# filtering out only non-null columns  
  
df4<-filter(df3, V1=="FALSE")  
df4

## Column\_Name V1  
## 1 YearStart FALSE  
## 2 YearEnd FALSE  
## 3 LocationAbbr FALSE  
## 4 LocationDesc FALSE  
## 5 DataSource FALSE  
## 6 Topic FALSE  
## 7 Question FALSE  
## 8 DataValueUnit FALSE  
## 9 DataValueType FALSE  
## 10 DataValue FALSE  
## 11 DataValueAlt FALSE  
## 12 DataValueFootnoteSymbol FALSE  
## 13 DatavalueFootnote FALSE  
## 14 LowConfidenceLimit FALSE  
## 15 HighConfidenceLimit FALSE  
## 16 StratificationCategory1 FALSE  
## 17 Stratification1 FALSE  
## 18 GeoLocation FALSE  
## 19 LocationID FALSE  
## 20 TopicID FALSE  
## 21 QuestionID FALSE  
## 22 DataValueTypeID FALSE  
## 23 StratificationCategoryID1 FALSE  
## 24 StratificationID1 FALSE

# creating a list of selected columns  
columns\_list<-df4$Column\_Name  
columns\_list

## [1] "YearStart" "YearEnd"   
## [3] "LocationAbbr" "LocationDesc"   
## [5] "DataSource" "Topic"   
## [7] "Question" "DataValueUnit"   
## [9] "DataValueType" "DataValue"   
## [11] "DataValueAlt" "DataValueFootnoteSymbol"   
## [13] "DatavalueFootnote" "LowConfidenceLimit"   
## [15] "HighConfidenceLimit" "StratificationCategory1"   
## [17] "Stratification1" "GeoLocation"   
## [19] "LocationID" "TopicID"   
## [21] "QuestionID" "DataValueTypeID"   
## [23] "StratificationCategoryID1" "StratificationID1"

#making a new dataframe with the selected non-null columns  
alcohol\_new<-subset(alcohol, select = columns\_list)  
subset\_alcohol\_new<-alcohol\_new[1:5, 1:6]  
kable(subset\_alcohol\_new)

| YearStart | YearEnd | LocationAbbr | LocationDesc | DataSource | Topic |
| --- | --- | --- | --- | --- | --- |
| 2021 | 2021 | AZ | Arizona | APIS | Alcohol |
| 2021 | 2021 | ME | Maine | Legal Research | Alcohol |
| 2021 | 2021 | VA | Virginia | APIS | Alcohol |
| 2021 | 2021 | MD | Maryland | APIS | Alcohol |
| 2021 | 2021 | WA | Washington | APIS | Alcohol |

colnames(alcohol\_new)

## [1] “YearStart” “YearEnd”   
## [3] “LocationAbbr” “LocationDesc”   
## [5] “DataSource” “Topic”   
## [7] “Question” “DataValueUnit”   
## [9] “DataValueType” “DataValue”   
## [11] “DataValueAlt” “DataValueFootnoteSymbol”   
## [13] “DatavalueFootnote” “LowConfidenceLimit”   
## [15] “HighConfidenceLimit” “StratificationCategory1”   
## [17] “Stratification1” “GeoLocation”   
## [19] “LocationID” “TopicID”   
## [21] “QuestionID” “DataValueTypeID”   
## [23] “StratificationCategoryID1” “StratificationID1”

# What are the unique survey questions in this dataframe  
unique(alcohol\_new$Question)

## [1] “Amount of alcohol excise tax by beverage type (distilled spirits)”   
## [2] “Commercial host (dram shop) liability laws”   
## [3] “Amount of alcohol excise tax by beverage type (beer)”   
## [4] “Amount of alcohol excise tax by beverage type (wine)”   
## [5] “Heavy drinking among adults aged >= 18 years”   
## [6] “Heavy drinking among women aged 18-44 years”   
## [7] “Binge drinking prevalence among adults aged >= 18 years”   
## [8] “Binge drinking intensity among adults aged >= 18 years who binge drink”  
## [9] “Binge drinking frequency among adults aged >= 18 years who binge drink”  
## [10] “Binge drinking prevalence among women aged 18-44 years”

# choosing only alcohol consumption related questions, total entries, 4130  
alcohol\_consumption<-subset(alcohol\_new, Question %in% c(“Heavy drinking among adults aged >= 18 years”,”Heavy drinking among women aged 18-44 years”,”Binge drinking prevalence among adults aged >= 18 years”,”Binge drinking intensity among adults aged >= 18 years who binge drink”,”Binge drinking frequency among adults aged >= 18 years who binge drink”,”Binge drinking prevalence among women aged 18-44 years”))  
  
unique\_questions<-unique(alcohol\_consumption$Question)  
unique\_questions

## [1] “Heavy drinking among adults aged >= 18 years”   
## [2] “Heavy drinking among women aged 18-44 years”   
## [3] “Binge drinking prevalence among adults aged >= 18 years”   
## [4] “Binge drinking intensity among adults aged >= 18 years who binge drink”  
## [5] “Binge drinking frequency among adults aged >= 18 years who binge drink”  
## [6] “Binge drinking prevalence among women aged 18-44 years”

# checking for missing values in rows and found 1,393 missing values  
missing\_rows<-alcohol\_consumption[apply(is.na(alcohol\_consumption), 1, any), ]  
nrow(missing\_rows)

## [1] 1393

# selecting rows without missing values, number of rows total, 2,737  
alcohol\_consumption\_new<-alcohol\_consumption[complete.cases(alcohol\_consumption), ]  
  
nrow(alcohol\_consumption\_new)

## [1] 2737

# Further inspecting for relevant rows  
columns\_list

## [1] “YearStart” “YearEnd”   
## [3] “LocationAbbr” “LocationDesc”   
## [5] “DataSource” “Topic”   
## [7] “Question” “DataValueUnit”   
## [9] “DataValueType” “DataValue”   
## [11] “DataValueAlt” “DataValueFootnoteSymbol”   
## [13] “DatavalueFootnote” “LowConfidenceLimit”   
## [15] “HighConfidenceLimit” “StratificationCategory1”   
## [17] “Stratification1” “GeoLocation”   
## [19] “LocationID” “TopicID”   
## [21] “QuestionID” “DataValueTypeID”   
## [23] “StratificationCategoryID1” “StratificationID1”

# selecting only relevant rows  
alcohol\_consumption\_mod<-subset(alcohol\_consumption\_new, select=c(“LocationDesc”,”Question”,”DataValueUnit”,”DataValueType”,”DataValueAlt”,”StratificationCategory1”,”Stratification1”))  
colnames(alcohol\_consumption\_mod)

## [1] “LocationDesc” “Question”   
## [3] “DataValueUnit” “DataValueType”   
## [5] “DataValueAlt” “StratificationCategory1”  
## [7] “Stratification1”

# SECTION 4:

## CHARACTERISTICS OF THE DATA:

# Description of columns in the final data set.  
data\_desc<-data.frame(Column\_Name = c(“LocationDesc”,”Question”,”DataValueUnit”,”DataValueType”,”DataValueAlt”,”StratificationCategory1”,”Stratification1”),  
Description = c(“US State Name”, “Alcohol survey question”, “Percentage”,”Prevalence/Age-adjusted Prevalence”, “Numerical value”, “Gender/Race”, “Male, Female, Black, Hispanic, Non-Hispanic, White”)  
 )  
column\_desc<-kable(data\_desc)  
column\_desc

| Column\_Name | Description |
| --- | --- |
| LocationDesc | US State Name |
| Question | Alcohol survey question |
| DataValueUnit | Percentage |
| DataValueType | Prevalence/Age-adjusted Prevalence |
| DataValueAlt | Numerical value |
| StratificationCategory1 | Gender/Race |
| Stratification1 | Male, Female, Black, Hispanic, Non-Hispanic, White |

The initial alcohol dataset had 4346 rows and 34 columns. After dropping null and irrelevant columns and dropping rows with missing entries, the cleaned dataset reduced to 7 columns and 2737 rows. The names of the columns are listed in the output above.

# 

# SECTION 5: SUMMARY STATISTICS ON HEAVY DRINKING ADULTS BASED ON GENDER AND RACE/ETHNICITY

# The final adult alcohol consumption dataset  
hdag\_18<-alcohol\_consumption\_mod %>% filter(Question=="Heavy drinking among adults aged >= 18 years")  
unique(hdag\_18$Question)

## [1] "Heavy drinking among adults aged >= 18 years"

# what are the data value types  
unique(hdag\_18$DataValueType)

## [1] "Crude Prevalence" "Age-adjusted Prevalence"

# filtering data set that gives adult male and female alcohol consumption data  
hdag\_18Gender<-hdag\_18 %>% filter(StratificationCategory1=="Gender")  
unique(hdag\_18Gender$StratificationCategory1)

## [1] "Gender"

# renaming Stratification1 as Gender  
  
names(hdag\_18Gender)[names(hdag\_18Gender) == 'Stratification1'] <- 'Gender'  
  
hdag\_18Gender$Gender[1:5]

## [1] "Male" "Female" "Male" "Male" "Female"

# segregating Crude Prevalence and Age Adjusted Prevalence  
filt1<-hdag\_18Gender %>% filter(DataValueType=="Crude Prevalence")  
filt2<-hdag\_18Gender %>% filter(DataValueType == "Age-adjusted Prevalence")

# assinging summary statistics to a new variable summrz\_g for prevalence of heavy drinking among adult males and females aged 18 and above   
summrz\_g<-filt1 %>% group\_by(Gender) %>% summarize(Mean\_Prevalence=mean(DataValueAlt), Min\_Prevalence=min(DataValueAlt),Max\_Prevalence=max(DataValueAlt), .groups= "keep")  
kable(summrz\_g)

| Gender | Mean\_Prevalence | Min\_Prevalence | Max\_Prevalence |
| --- | --- | --- | --- |
| Female | 5.967924 | 2.6 | 11.1 |
| Male | 6.568518 | 4.5 | 9.2 |

# assigning summary statistics to a new variable summrz\_g2, for age adjusted prevalence among adult males and females aged 18 and above   
summrz\_g2<-filt2 %>% group\_by(Gender) %>% summarize(Mean\_Prevalence=mean(DataValueAlt), Min\_Prevalence=min(DataValueAlt),Max\_Prevelance=max(DataValueAlt), .groups= "keep")  
kable(summrz\_g2)

| Gender | Mean\_Prevalence | Min\_Prevalence | Max\_Prevelance |
| --- | --- | --- | --- |
| Female | 6.216981 | 2.7 | 10.2 |
| Male | 6.746296 | 4.5 | 9.7 |
|  |  |  |  |

# filtering alcohol consumption among different Race/Ethinicty  
hdag\_18Race<-hdag\_18 %>% filter(StratificationCategory1=="Race/Ethnicity")  
unique(hdag\_18Race$StratificationCategory1)

## [1] "Race/Ethnicity"

# Rename Stratification 1 as Race/Ethnicity  
  
names(hdag\_18Race)[names(hdag\_18Race) == 'Stratification1'] <- 'Race\_Ethnicity'  
  
hdag\_18Race$Race\_Ethnicity[1:5]

## [1] "Multiracial, non-Hispanic" "Black, non-Hispanic"   
## [3] "White, non-Hispanic" "White, non-Hispanic"   
## [5] "Black, non-Hispanic"

# checking for DataValueType  
unique(hdag\_18Race$DataValueType)

## [1] "Age-adjusted Prevalence" "Crude Prevalence"

# Segregating cure prevalence from age-adjusted prevalence  
filt3<-hdag\_18Race %>% filter(DataValueType=="Crude Prevalence")  
filt4<-hdag\_18Race %>% filter(DataValueType == "Age-adjusted Prevalence")

# assigning summary statistics to a new variable summrz\_g3 for prevalence of heavy alcohol consumption among different race/ethnicity aged 18 and above   
summrz\_g3<-filt3 %>% group\_by(Race\_Ethnicity) %>% summarize(Mean\_Prevalence=mean(DataValueAlt), Min\_Prevalence=min(DataValueAlt),Max\_Prevalence=max(DataValueAlt), .groups= "keep")  
kable(summrz\_g3)

| Race\_Ethnicity | Mean\_Prevalence | Min\_Prevalence | Max\_Prevalence |
| --- | --- | --- | --- |
| Black, non-Hispanic | 5.309091 | 2.7 | 10.0 |
| Hispanic | 4.782143 | 2.4 | 8.3 |
| Multiracial, non-Hispanic | 8.483333 | 2.8 | 13.3 |
| Other, non-Hispanic | 4.835714 | 2.2 | 7.2 |
| White, non-Hispanic | 6.910000 | 3.8 | 11.7 |

# assigning summary statistics to a new variable summrz\_g4, for age-adjusted prevalence of heavy drinking among different race/ethinicity aged 18 and above  
summrz\_g4<-filt4 %>% group\_by(Race\_Ethnicity) %>% summarize(Mean\_Prevalence=mean(DataValueAlt), Min\_Prevelence=min(DataValueAlt),Max\_Prevelance=max(DataValueAlt), .groups= "keep")  
kable(summrz\_g4)

| Race\_Ethnicity | Mean\_Prevalence | Min\_Prevelence | Max\_Prevelance |
| --- | --- | --- | --- |
| Black, non-Hispanic | 5.617391 | 3.3 | 10.1 |
| Hispanic | 4.582759 | 2.2 | 8.2 |
| Multiracial, non-Hispanic | 8.571429 | 2.8 | 13.6 |
| Other, non-Hispanic | 4.500000 | 1.8 | 6.9 |
| White, non-Hispanic | 7.513725 | 4.0 | 17.3 |

## CONCLUSION:

From the summary statistics obtained from summrz\_g, summrz\_g2, summrz\_g3 and summrz\_g4 (variables assigned for summary statistics for crude prevalence and age-adjusted prevalence based on gender or race/ethnicity) the following results are obtained. The mean crude prevalence rate of heavy drinking among men (6.5685185%) is higher than the mean crude prevalence rate among women (5.9679245%). Even after adjusting for age, the mean prevalence rate for men (6.7462963%) is still higher than women (6.2169811%).

The mean crude prevalence rate of heavy drinking among multiracial, Hispanic(8.4833333%) is the highest followed by White, non-Hispanic(6.91%). Adjusting for age, did not significantly change the outcome for heavy drinking rate.

Note: The above conclusions are highly contingent on the available data and is subject to change anytime with new available data. This is solely meant for academic exercise purpose only and in no way represents the personal view of the author nor it is intended to be discriminatory or maliciously targeted towards any particular race/ethnicity or gender.