The Worlds Alcohol Consumption Database for machine learning National Alcohol Surveys

ARG and its Center have conducted a series of National Alcohol Surveys (NAS) of the adult (age 18 and older) US population at approximately 5-year intervals since the 1960s with considerable standardization of measurement methods since 1979 (the 6th NAS or N6). Although not unique in conducting repeated surveys including alcohol questions, the NAS has the longest time frame with commensurate, detailed alcohol use pattern and problem measures, which now extends over 35 years covered by 8 surveys (N6 TO N13), with the most recent survey conducted during 2014 and early 2015.

NAS datasets and codebooks for N6 through N12 (2010) are <u>available from ARG</u>. Requests for N13 data will be evaluated on a case-by-case basis, as analyses using this survey are ongoing under various current grants to ARG Scientists. For more information, please contact Study Director <u>Dr. Kate Karriker-Jaffe</u>.

National 13, 2014-2015 (N13) 7,071 Random Digit Dial (RDD) nationally representative Computer Assisted Telephone Interview (CATI, of landline (n=4,109) and cellular (n=2,962) phones) of adults aged 18+ in 50 states (plus Washington, DC), with oversamples of Black non-Hispanic (total n = 1,763) and Hispanic (total n = 1,623) groups (with interviews in Spanish for those requesting/needing this). Sample contains 5,634 complete cases and 991 partially-complete cases (defined as those who completed demographics, alcohol consumption patterns, alcohol problems, and alcohol treatment items).

These N13 data are currently being analyzed under aims in the National Alcohol Research Center (P50 AA005595). Use by or collaborations with other researchers would be based on discussions with the Center PI (William Kerr) and under data use agreements developed with Project Co-Directors, Dr. Katherine Karriker-Jaffe & Thomas Greenfield. <u>N13 Questionnaire</u>

2015 National Alcohol's Harm to Others Survey (NAHTOS)

The 2015 NAHTOS (R01AA022791, M-PI T. Greenfield and K. Karriker-Jaffe) is a telephone survey that used the same sampling strategy as N13, collecting data from 2,591 cases (2,440 complete interviews) to assess types, sources and severity of alcohol's harm to others (1,763 landline and 1,945 cellular phone cases). Fieldwork was completed in the Spring of 2015. The survey instrument contains indicators of alcohol and other drug use, mental health (anxiety and

depression) and perceptions of the neighborhood/community context, including social cohesion, crime and disorder. Added to these are extensive assessments of harms from others' drinking (intimate partners/family members/friends and strangers), with some items also assessed on N13.

Alcohol Consumed By Country

Overview

This data set contains data on alcohol consumption per capita across countries in the world.

Details

World   alcohol   consumption   dataset
This is a global beverage consumption record dataset. The first column means the year of the record, the second column refers to the place where the beverage was produced, and the third column refers to the place where the beverage was consumed. The fourth columns refer to the types of beverages, and the fifth column refers to the average consumption of beverages per person.

It contains data on average alcohol consumption by country among those 15 and older, separately by beverage type. Values provided are the average number of servings in three categories (beer, wine, and spirits servings).

Data Description

Variable	Description
Year	Survey of year
WHO Region	Country comes under WHO Region
Country	Name of Country
Beverage Types	Type of beverage like Spirits, Wine, Beer or Other
Display Value	Value of consuming beverage for respective country in WHO Region

```
alcohol_data = read.csv("data/world_alcohol_consumption.csv")
alcohol_data = alcohol_data[, 2:5]
head(alcohol_data)
```

	Country	beer_servings	spirit_servings
	r>	:>	:>
1	Afghanistan	0	0
2	Albania	89	132
3	Algeria	25	0
4	Andorra	245	138
5	Angola	217	57
6	Antigua & Barbuda	102	128

Data Files

- world_alcohol_consumption.csv

Objectives

We will consider modeling the average consumption of beer, wine, and spirits across countries. The goal is to find distributions that fit these data well, and to estimate the associated parameters, as well as to produce maps displaying information on alcohol consumption across countries.

- Provide numerical and graphical summaries of the average number of beer, wine, and spirit servings per capita per country.
- Consider fitting some of the known distributions discussed in class, both this semester and last semester, to these variables. For each distribution you consider, explain how you are estimating the relevant parameters (e.g., are you using the MLEs? MOMs? etc.). Consider at least two different distributions for each of the variables.
- Provide some assessment of fit of the distributions to the data. For example, q-q plots are useful for this purpose.

Project Description

Summary:

Makes columns showing the number of beers, glasses of wine, shots of liquor, and total drinks consumed based on the amount of ethanol consumed for each category that was already included.

This data set contains the per capita (persons aged 14+) consumption of ethanol (in gallons) for each state, Washington D.C., and totals for census regions and the United States as a whole, for the years 1977-2018. This includes total ethanol consumed as well as consumption by three categories: beer, wine, and shots of liquor ("spirits"). The PDF includes a method to convert the ethanol variables into total drinks of each type. I used this method to create columns for how many beers, glasses of wine, shots of liquor, and total drinks were consumed.

The PDF doesn't say how many ounces of fluid is in each drink type (except for the *number_of_drinks_total* variable) so I used the information provided by the National Institute on Alcohol Abuse and Alcoholism here - https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/what-standard-drink. Please note that the *number_of_drinks_total* variable is based on the conversion formula provided, not by adding the individual drink categories together and therefore will be slightly different than that way of measuring it.

This data comes from a report by Dr. Megan E. Slater and Dr. Hillel R. Alpert at the National Institute on Alcohol Abuse and Alcoholism (downloaded here https://pubs.niaaa.nih.gov/publications/surveillance115/CONS18.htm). That report is one of the files available to download and is included as it explains the methodology the two authors used for the data. I am not affiliated with the original report at all. If you do use this data please also cite the original report.

When using this data consider that it is rate per capita (persons aged 14+) based on the population in that state so states that experience lots of visitors (e.g. Nevada and Washington D.C.) may have incorrect numbers.

Scope of Project

Subject Terms: Alcohol consumption; alcohol; drinking behavior; beer; liquor; spirits; wine; booze; alcohol use; NIH

Geographic Coverage: United States

Time Period(s): 1977 – 2018

Universe: Persons aged 14 and over living in the United States

Data Type(s): Aggregate data

Use Cases:

1. Write a Python program to find out the alcohol consumption details in the year '1987' or '1989' from the dataset.

2. Write a python program to find out and display horizontal bar chart the alcohol consumption details in the year '1986' where WHO region is 'Western Pacific' and country is 'VietNam' from thedataset.

3. Write a Pandas program to find and display bar chart to out the alcohol consumption details in the year '1986' or '1989' where WHO region is 'Americas' or 'Europe' from the world alcohol consumption dataset. Use Different colors for each bar and attach text label above each bar.

4. Write a Pandas program to find out the 'WHO region, 'Country', 'Beverage Types' in the year '1986' or '1989' where WHO region is 'Americas' or 'Europe' from the world alcohol consumption dataset.

5. Write a Pandas program to find out the records where consumption of beverages per person average >=5 and Beverage Types is Beer from world alcohol consumption dataset.

6. Write a Pandas program to find out and display Pie chart of the records where consumption of beverages per person average >=4 and Beverage Types is Beer, Wine, Spirits from world alcohol consumption dataset.

7. Write a Pandas program to filter the WHO region and Beverage Types columns and records by range from 0 to 15 from dataset.

8. Write a Pandas program to filter those records where WHO region contains "Ea" substring from dataset and show count of it.

9. Write a Pandas program to filter those records where WHO region matches with multiple values (Africa, Eastern Mediterranean, Europe) from dataset.

10. Write a Pandas program to filter those records which not appears in a given list from dataset.

    who_region = ["Africa", "Eastern Mediterranean", "Europe"]

11. Write a Pandas program to create bar plots with error bars to filter all records where the average consumption of beverages per person from 0.5 to 2.50 in dataset.

12. Write a Pandas program to find average consumption of wine per person greater than 2 in world alcohol consumption dataset.

13. Write a Pandas program to filter rows, based on row numbers ended with 0, like 0, 10, 20, 30 from dataset.

14. Write a Pandas program to also select rows with Index label 0 to 9 with some columns from dataset.

15. Write a Pandas program to rename all and only some of the column names from dataset by adding _ in column names.

16. Write a Pandas program to find which years have all non-zero values and which years have any non-zero values from dataset.

17.   Write a Pandas program to filter all columns where all entries present, check which rows and columns has a NaN and finally drop rows with any NaNs from world alcohol consumption dataset.

18.   Write a Pandas program to filter all records starting from the 'Year' column, access every other column from world alcohol consumption dataset.

19.   Write a Pandas program to filter all records starting from the 2nd row, access every 5th row from dataset.

Thank You !

Editor: Rajesh Gawai

Research Unit:

Published Online : 31 May 2021