World Alcohol Consumption:

**Who drinks the most, and where?**

**Mari Osmundson**

*Dataset info and discussion:*

I found a dataset on fivethirtyeight containing data on alcoholic beverage consumption for the world: <https://github.com/fivethirtyeight/data/tree/master/alcohol-consumption>. I decided to use this dataframe as a basis for an expanded analysis. I was curious if there was a correlation between extreme temperatures and increased alcohol consumption. I also wanted to see which world regions drank the most alcohol.

The world temperature data was obtained from <https://www.revolvy.com/page/List-of-countries-by-average-yearly-temperature>. The data reflects average temperatures from 1961-1990.

Next I created the World\_Region csv file with the help of www.worldatlas.com and Pew Research Center. There was great variation on the Internet for divvying up the world into regions. For the purposes of this analysis, I split the world’s countries into eight regions: Africa, Asia, Caribbean, Europe, Middle East, North America, Oceania, and South America. I chose to combine Central America with Canada and the USA for the sake of comparison.

After some analysis, I became curious to see how majority belief system might be connected to alcohol consumption. I added data in Excel to the drinks\_join dataframe, then re-import the dataframe into RStudio. The belief system data was obtained from the Pew Research Center Global Religious Futures project: <http://www.globalreligiousfutures.org>.

Conclusions: There was a weak negative correlation between average temperature and average servings. Regions which observed some form of Christianity as the majority belief system also had the highest number of average servings. Asia has the greatest belief system diversity; almost all major religions are observed on that continent, with the exception being Judaism.

*Data cleaning/analysis in R:*

I changed the country column in drinks and World\_Region for it to be used as the key in a full-join later.

drinks <- drinks %>%  
 rename(Country = country)  
  
World\_Region <- World\_Region %>%  
 rename(Country = country)

I opted for a full join to retain all data from each dataframe.

drinks\_join <- drinks %>%  
 full\_join(avg\_temps\_of\_the\_world, by = "Country") %>%  
 full\_join(World\_Region, by = "Country")

I cleaned the data. Each dataframe had some slight differences in naming conventions (& versus ‘and’, hyphen versus no hyphen), which caused the data to not merge in the join.

#Antigua & Barbuda  
drinks\_join[6, 6] = 26.00  
  
drinks\_join <- drinks\_join[-194, ]

#Bosnia-Herzegovina   
drinks\_join[22, 6] = 9.85  
  
drinks\_join <- drinks\_join[-194, ]

#Cote dIvoire   
drinks\_join[29, 6] = 26.35  
  
drinks\_join <- drinks\_join[-195, ]

#Macedonia   
drinks\_join[171, 6] = 9.8  
  
drinks\_join <- drinks\_join[-198, ]

#Russia  
drinks\_join[199, 2:7] = c(247, 326, 73, 11.5, -5.10, "Asia")  
  
drinks\_join <- drinks\_join[-142, ]

#Micronesia   
drinks\_join[111, 6] = 25.85  
  
drinks\_join <- drinks\_join[-195, ]

#Sao Tome   
drinks\_join[148, 6] = 23.75  
  
drinks\_join <- drinks\_join[-201, ]

#Trinidad and Tobago   
drinks\_join[174, 6] = 25.75  
  
drinks\_join <- drinks\_join[-201, ]

#Eswatini (Swaziland)   
drinks\_join[202, 2:6] = c(90, 2, 2, 4.7, 21.4)  
  
drinks\_join <- drinks\_join[-194, ]  
  
drinks\_join <- drinks\_join[-164, ]

#United States   
drinks\_join[183, 6] = 8.55  
  
drinks\_join <- drinks\_join[-199, ]

#Cabo Verde   
drinks\_join[30, 6] = 23.30  
  
drinks\_join <- drinks\_join[-192, ]

write.csv(drinks\_join, file = "drinks\_join.csv")

I changed C. America to N. America in the Region column.

Regions\_sub <- gsub("C. America", "N. America", drinks\_join$Region)

view(Regions\_sub)

drinks\_join <- drinks\_join %>%  
 mutate(Region = Regions\_sub)

Added region to rows 90 and 183

drinks\_join[183, 7] <- "N. America"  
  
drinks\_join[40, 7] <- "Africa"

Changed column types to numeric for future calculations.

drinks\_join <- drinks\_join %>%  
 mutate\_at(2:6, as.numeric)

drinks\_join <- drinks\_join %>%  
 mutate\_at(8, as.numeric)

Calculated average servings of alcohol for each country and added it to the dataframe as a column.

drinks\_join <- drinks\_join %>%  
 mutate(avg\_servings = (`beer\_servings` + `spirit\_servings` + `wine\_servings`) / 3)

Created a vector to round average servings.

drinks\_rounded <-   
 signif(drinks\_join$avg\_servings, digits = 2)

Overwrote the data in avg\_servings with the data from drinks\_rounded.

drinks\_join <- drinks\_join %>%  
 mutate(avg\_servings = drinks\_rounded)

Ran a correlation for average servings vs average temperature.

correlation\_matrix <- drinks\_join %>%  
 select(avg\_servings, `Average yearly temperature (1961–1990, degrees Celsius)`) %>%  
 cor(use = "pairwise.complete.obs")

Gathered the data by servings.

drinks\_gather <- drinks\_join %>%  
 gather(key = "serving", value = 'No. of servings', beer\_servings:wine\_servings) %>%  
 drop\_na()

Grouped the data by region.

drinks\_group <- drinks\_join %>%  
 drop\_na(avg\_servings) %>%  
 group\_by(Region) %>%  
 summarize(avg\_srvg\_region = mean(avg\_servings, na.rm = TRUE),  
 avg\_temp\_region = mean(`Average yearly temperature (1961–1990, degrees Celsius)`, na.rm = TRUE))

Here I added the majority belief system to the grouped dataframe, obtained from the Pew Research Center.

drinks\_group <- drinks\_group %>%  
 mutate(region\_top\_religion = c("Christianity", "Hinduism", "Christianity", "Christianity", "Islam", "Christianity", "Christianity", "Christianity"))

drinks\_group <- drinks\_group %>%  
 rename(region\_main\_religion = region\_top\_religion)

drinks\_group <- drinks\_group[ ,-5]

*The following pages describe plots made in RStudio and Tableau, and the conclusions drawn from those plots.*

I created a boxplot in RStudio based on regional averages in temperature and servings, reflecting the decrease in serving size as the average temperature rises.

ggplot(data = drinks\_join) +  
 aes(x = 'Average yearly temperature (1961-1990, degrees Celsius)', y = avg\_servings, fill = Region) +  
 geom\_boxplot()



I created a histogram for average servings. Average servings in the zero-50 range were more numerous than higher numbers of servings.

avg\_servings\_histo <- ggplot(data = drinks\_join) +  
 aes(x = avg\_servings) +  
 geom\_histogram(binwidth = 20)





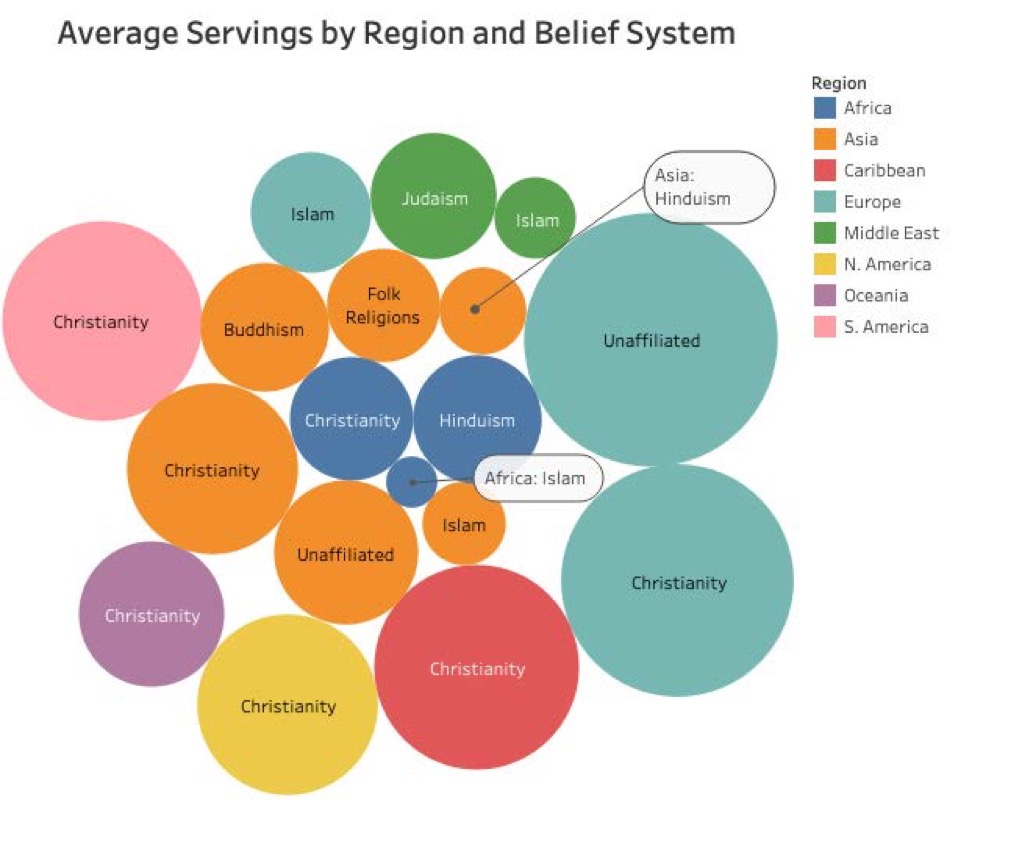
This Tableau plot reflects the weak negative correlation between temperature and average serving size. As the temperature increased, the number of average servings consumed decreased.

A close up of a logo

Description automatically generated

As shown in the Tableau plot to the left, Christianity had significantly more servings than other belief systems.

The Tableau plot to the right shows the beliefs diversity in Asia; the most religiously diverse continent in the world.



**Summary of Conclusions**

* There is a weak negative correlation between average servings and average temperature around the globe.
* Asia is the most religiously and spiritually diverse region in the world.
* Christianity had significantly more alcohol consumption than other belief systems: a fact worthy of note to those companies involved in alcohol sales and marketing.