

## ETL Project Report

Big Ballers: Yannik Alinyoh, Tara Flowers, Haochen (Harry) Zhang, Zhuowen (Ellie) Zhang

Georgia Tech Data Analysis BootCamp

03/10/2020

First, we extracted the raw datasets directly from the CDC and America's Health Rankings' websites. In particular, we exported CSV formats of three datasets on alcohol consumption, excessive drinking, and binge drinking.

Second, we transformed the datasets by filtering columns (variables) that we consider relevant. We then re-named the variables to make it easily understandable in further analyses.

Third, we chose these datasets based on their ability to be joined by state. The output database presents the relevance of alcohol consumption based on state by year.

## Reference:

“BRFSS Prevalence & Trends Data: Explore by Topic.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention,  
nccd.cdc.gov/BRFSSPrevalence/rdPage.aspx?rdReport=DPH\_BRFSS.ExploreByTopic&irbLocationType=StatesAndMMSA&islClass=CLASS01&islTopic=TOPIC03&islYear=2018&rdRnd=28685.

“BRFSS: Table of Alcohol Consumption.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention,  
chronicdata.cdc.gov/Behavioral-Risk-Factors/BRFSS-Table-of-Alcohol-Consumption/dts9-xy2f.

“Public Health Impact: Excessive Drinking.” *America's Health Rankings*, America's Health Rankings,  
[www.americashealthrankings.org/explore/annual/measure/ExcessDrink/state/ALL?ed](http://www.americashealthrankings.org/explore/annual/measure/ExcessDrink/state/ALL?edition-year=2016)

[i](#)

tion-year=2016.