Comparison of Alcohol Use by Youth in 2013 and 2015

Ellen Chancey

Primary Analysis Objectives

Determine whether the rate of alcohol use by youth, as reported by the Centers for Disease Control and Prevention (CDC) in their US Chronic Disease Indicators data, is significantly lower in 2015 than in 2013.

Background

The CDC has identified 124 indicators of chronic disease, one of which is the rate of alcohol use by youth. The CDC collects and distributes data on these indicators for use in public health efforts.

Data Sources

The following dataset will be used throughout the statistical analyses:

US Chronic Disease Indicators: https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi-e50c9

Data regarding use of alcohol by youth is extracted from the dataset, which contains all 124 indicators, and is then narrowed to only those states and territories that reported data in both 2013 and 2015. Finally, the data is formatted into two different structures. The difference between the two structures is that the youthalc1 dataset contains a record for each state or territory for each year. The second dataset, youthalc2 contains one record for each state or territory, with the datapoint for each year as a variable within the dataset.

youthalc1

This dataset is used to conveniently plot data before analysis.

```
display_output(youthalc1, out_type)
```

ï..LocationAbbr	LocationDesc	Rate	Year
AK	Alaska	22.0	2015
AL	Alabama	30.7	2015
AR	Arkansas	27.6	2015
AZ	Arizona	34.8	2015
CT	Connecticut	30.2	2015
DC	District of Columbia	20.1	2015
FL	Florida	33.0	2015
GU	Guam	25.3	2015
HI	Hawaii	25.1	2015
ID	Idaho	28.3	2015
IL	Illinois	30.7	2015
KY	Kentucky	28.5	2015
MA	Massachusetts	33.9	2015
MD	Maryland	26.1	2015
ME	Maine	24.0	2015
MI	Michigan	25.9	2015
MO	Missouri	34.5	2015
MS	Mississippi	31.5	2015

ï..LocationAbbr	LocationDesc	Rate	Year
MT	Montana	34.2	2015
NC	North Carolina	29.2	2015
ND	North Dakota	30.8	2015
NE	Nebraska	22.7	2015
NH	New Hampshire	29.9	2015
NM	New Mexico	26.1	2015
NV	Nevada	33.5	2015
NY	New York	29.7	2015
OK	Oklahoma	27.3	2015
PR	Puerto Rico	21.2	2015
RI	Rhode Island	26.1	2015
SC	South Carolina	24.6	2015
SD	South Dakota	28.0	2015
VA	Virginia	23.4	2015
WV	West Virginia	31.1	2015
WY	Wyoming	31.0	2015
AK	Alaska	22.5	2013
AL	Alabama	35.0	2013
AR	Arkansas	36.3	2013
AZ	Arizona	36.0	2013
CT	Connecticut	36.7	2013
DC	District of Columbia	31.4	2013
FL	Florida	34.8	2013
GU	Guam	23.5	2013
HI	Hawaii	25.2	2013
ID	Idaho	28.3	2013
IL	Illinois	36.6	2013
KY	Kentucky	30.4	2013
MA	Massachusetts	35.6	2013
MD	Maryland	31.2	2013
ME	Maine	26.6	2013
MI	Michigan	28.3	2013
MO	Missouri	35.6	2013
MS	Mississippi	32.9	2013
MT	Montana	37.1	2013
NC	North Carolina	32.2	2013
ND	North Dakota	35.3	2013
NE	Nebraska	22.1	2013
NH	New Hampshire	32.9	2013
NM	New Mexico	28.9	2013
NV	Nevada	34.0	2013
NY	New York	32.5	2013
OK	Oklahoma	33.4	2013
PR	Puerto Rico	25.5	2013
RI	Rhode Island	30.9	2013
SC	South Carolina	28.9	2013
SD	South Dakota	30.8	2013
VA	Virginia	27.3	2013
WV	West Virginia	37.1	2013
WY	Wyoming	34.4	2013

youthalc2

This dataset is used to complete analysis.

```
display_output(youthalc2, out_type)
```

ï..LocationAbbr	LocationDesc	Rate2015	Rate2013
AK	Alaska	22.0	22.5
AL	Alabama	30.7	35.0
AR	Arkansas	27.6	36.3
AZ	Arizona	34.8	36.0
CT	Connecticut	30.2	36.7
DC	District of Columbia	20.1	31.4
FL	Florida	33.0	34.8
GU	Guam	25.3	23.5
HI	Hawaii	25.1	25.2
ID	Idaho	28.3	28.3
IL	Illinois	30.7	36.6
KY	Kentucky	28.5	30.4
MA	Massachusetts	33.9	35.6
MD	Maryland	26.1	31.2
ME	Maine	24.0	26.6
MI	Michigan	25.9	28.3
MO	Missouri	34.5	35.6
MS	Mississippi	31.5	32.9
MT	Montana	34.2	37.1
NC	North Carolina	29.2	32.2
ND	North Dakota	30.8	35.3
NE	Nebraska	22.7	22.1
NH	New Hampshire	29.9	32.9
NM	New Mexico	26.1	28.9
NV	Nevada	33.5	34.0
NY	New York	29.7	32.5
OK	Oklahoma	27.3	33.4
PR	Puerto Rico	21.2	25.5
RI	Rhode Island	26.1	30.9
SC	South Carolina	24.6	28.9
SD	South Dakota	28.0	30.8
VA	Virginia	23.4	27.3
WV	West Virginia	31.1	37.1
WY	Wyoming	31.0	34.4

Analysis Methods

About the Data

Mean and Standard Deviation

Mean and standard deviation provide a general picture of the data to compare the two years before analysis.

2013

```
mean(youthalc2$Rate2013)
```

```
## [1] 31.47647
```

```
sd(youthalc2$Rate2013)
```

```
## [1] 4.406542
```

2015

```
mean(youthalc2$Rate2015)
```

```
## [1] 28.26471
```

```
sd(youthalc2$Rate2015)
```

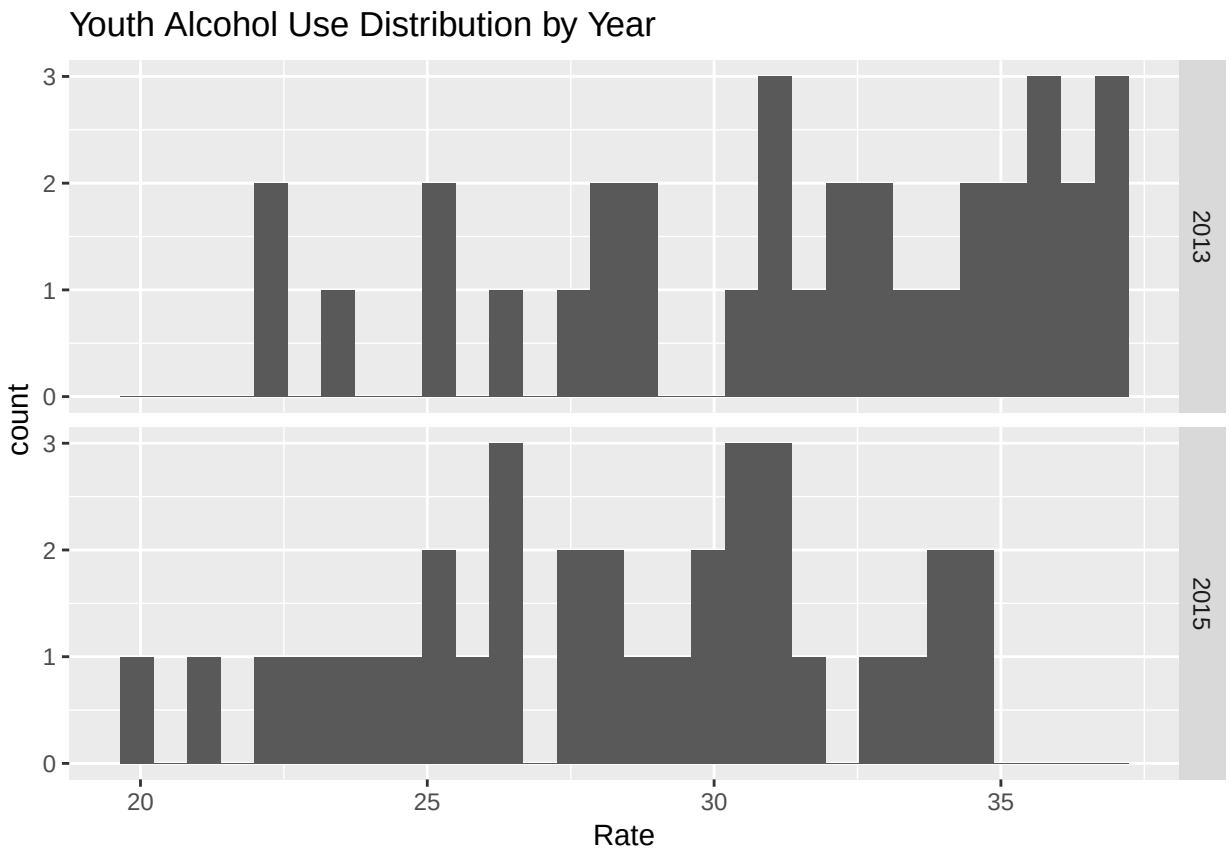
```
## [1] 4.02288
```

Basic plots

Distribution of Youth Alcohol Use by Year

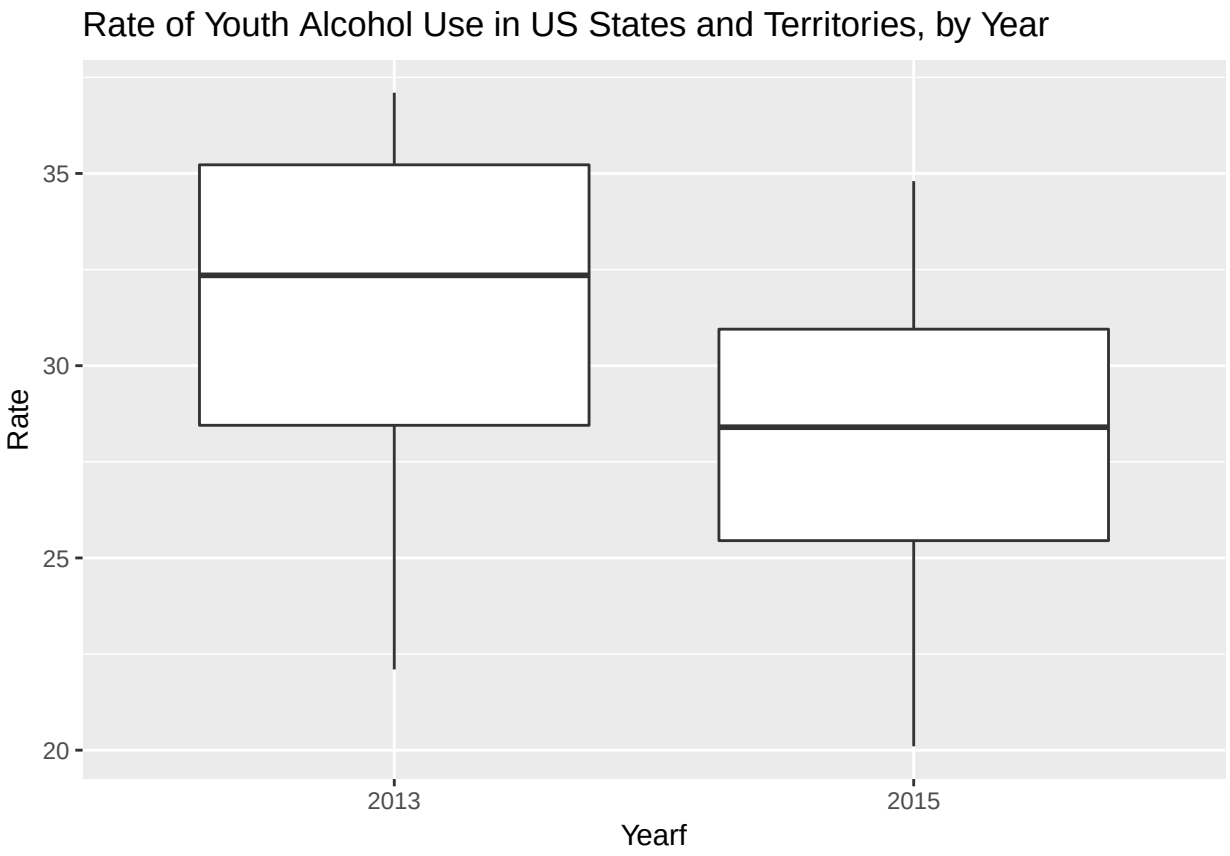
```
ggplot(data=youthalc1, aes(x=Rate))+
  geom_histogram() +
  facet_grid(Year~.) +
  ggtitle("Youth Alcohol Use Distribution by Year")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Boxplot of Youth Alcohol Use by Year

```r
Yearf<-factor(youthalc1$Year)
ggplot(data=youthalc1, aes(x=Yearf,y=Rate))+
  geom_boxplot() +
  ggtitle("Rate of Youth Alcohol Use in US States and Territories, by Year")
```



Assumptions

All inferences are conducted using $\alpha = 0.05$ unless stated otherwise. No adjustments for multiplicity are made as this is an exploratory analysis. Discrete variables are summarized with proportions and frequencies. Continuous variables are summarized using the following statistics:

- mean
- median
- standard deviation
- coefficient of variation
- quantiles
- minimum
- maximum

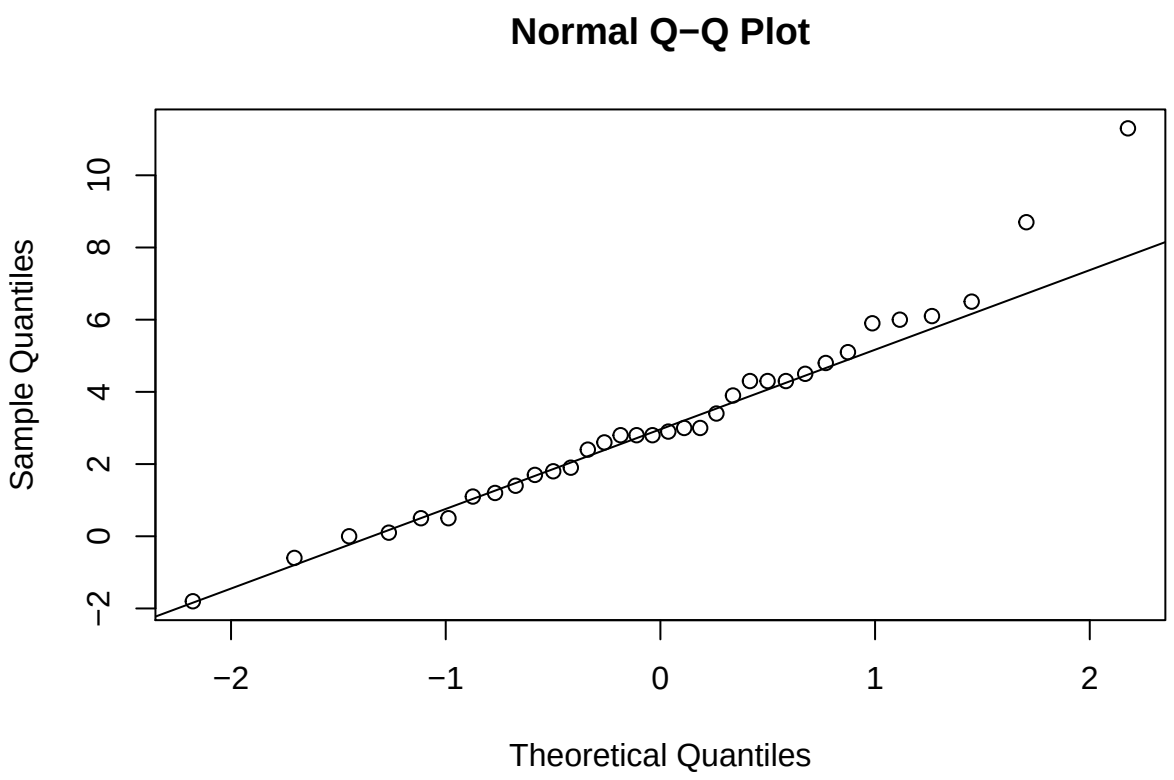
Normality of Paired Differences

Normality of paired differences is tested by use of a QQ Plot and the Shapiro-Wilk test.

QQ Plot

According to the QQ Plot, the years have similar distributions.

```
difference<-youthalc2$Rate2013-youthalc2$Rate2015
qqnorm(difference)
qqline(difference)
```



Shapiro-Wilk Test

According to the Shapiro-Wilk test there is insufficient evidence to reject the null hypothesis of normality (p=0.232)

```
shapiro.test(difference)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  difference
## W = 0.9593, p-value = 0.232
```

Primary Objective Analysis

The primary objective analysis uses a paired t test to determine if there is a significant difference between alcohol use by youth in 2013 and 2015.

Primary Objective Results

```
t.test(youthalc2$Rate2013, youthalc2$Rate2015, paired=TRUE, alternative="g")
```

```
##
##  Paired t-test
##
## data:  youthalc2$Rate2013 and youthalc2$Rate2015
## t = 7.0123, df = 33, p-value = 2.537e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.436631      Inf
## sample estimates:
## mean of the differences
##                3.211765
```

Conclusions and Discussion

There is sufficient evidence to conclude that the rate of alcohol consumption by youth was higher in 2013 than in 2015 (p = 2.537e-08). On average the rate in 2013 was 2.44 points higher than in 2015 (95% CI).

One limitation of these findings is that only states and territories that reported data in both 2013 and 2015 were included. A more complete picture of the federal trend could be determined if all states and territories had both data points.

Information

All of the statistical analyses in this document will be performed using R version 3.4.1 (2017-06-30). R packages used will be maintained using the packrat dependency management system.

```
sessionInfo()
```

```
## R version 3.4.1 (2017-06-30)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 14393)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] knitr_1.16    DT_0.2        car_2.1-5     dplyr_0.7.2   ggplot2_2.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.11       highr_0.6          nloptr_1.0.4
##  [4] compiler_3.4.1     plyr_1.8.4         bindr_0.1
```

```
##  [7] tools_3.4.1         digest_0.6.12       lme4_1.1-13
## [10] evaluate_0.10.1     tibble_1.3.3        gtable_0.2.0
## [13] nlme_3.1-131        lattice_0.20-35     mgcv_1.8-17
## [16] pkgconfig_2.0.1     rlang_0.1.1         Matrix_1.2-10
## [19] parallel_3.4.1      yaml_2.1.14         SparseM_1.77
## [22] bindrcpp_0.2        stringr_1.2.0       htmlwidgets_0.9
## [25] MatrixModels_0.4-1 rprojroot_1.2       grid_3.4.1
## [28] nnet_7.3-12         glue_1.1.1          R6_2.2.2
## [31] rmarkdown_1.6       minqa_1.2.4         reshape2_1.4.2
## [34] magrittr_1.5        backports_1.1.0     scales_0.4.1
## [37] htmltools_0.3.6     MASS_7.3-47         splines_3.4.1
## [40] assertthat_0.2.0    pbkrtest_0.4-7      colorspace_1.3-2
## [43] labeling_0.3        quantreg_5.33       stringi_1.1.5
## [46] lazyeval_0.2.0      munsell_0.4.3
```