

# Happines,Soccer, Alcohol, and Trump (Milestone2)

*Alexander Chang*

*March 27, 2018*

```
library(printr)

## Warning: package 'printr' was built under R version 3.3.3
library(gplots)

## Warning: package 'gplots' was built under R version 3.3.3
##
## Attaching package: 'gplots'
## The following object is masked from 'package:stats':
##
##      lowess
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.3
load("C:/Users/Alex Chang/Documents/GitHub/BigDataProject/mydata.RData")
library(Hmisc)

## Warning: package 'Hmisc' was built under R version 3.3.3
## Loading required package: lattice
## Loading required package: survival
## Warning: package 'survival' was built under R version 3.3.3
## Loading required package: Formula
## Warning: package 'Formula' was built under R version 3.3.3
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##      format.pval, units
```

## Github Respository: <https://github.com/theamazingchang/BigDataProject>

It's very common to enjoy a drink every now and then. Some people more than others, and some countries more than others. Today I will perform some fun statistics on the combination of four different data sets that observes happiness, alcohol consumption, fifa (soccer) participation, and their approval of our esteemed leader, President Trump.

We begin by performing a merge of all the datasets into one big dataset

```
merge1 <- merge(happy, drinks, by="Country")
merge2 <- merge(merge1, fifa, by="Country", all = FALSE)
TrumpMerge <- merge(merge2, Trump, by = "Country", all = TRUE)
```

Let's observe if there are correlations hidden within this database by producing a correlation matrix

```
corr1 <- rcorr(as.matrix(mergeframe))
corr1
```

```
##              Happiness.Rank Happiness.Score Whisker.high
## Happiness.Rank              1.00          -0.99         -0.99
## Happiness.Score            -0.99              1.00          1.00
## Whisker.high              -0.99              1.00          1.00
## Whisker.low              -0.99              1.00          1.00
## Economy..GDP.per.Capita.  -0.81              0.81          0.81
## Family                    -0.74              0.75          0.75
## Health..Life.Expectancy. -0.78              0.79          0.78
## Freedom                  -0.54              0.56          0.56
## Generosity               -0.12              0.14          0.14
## Trust..Government.Corrup. -0.42              0.44          0.44
## Dystopia.Residual        -0.52              0.51          0.51
## beer_servings            -0.51              0.51          0.51
## spirit_servings           -0.31              0.29          0.29
## wine_servings            -0.49              0.50          0.49
## total_litres_of_pure_alcohol -0.46              0.45          0.45
## population_share          0.04             -0.04         -0.05
## tv_audience_share        -0.15              0.14          0.13
## gdp_weighted_share        -0.32              0.30          0.29
##              Whisker.low Economy..GDP.per.Capita. Family
## Happiness.Rank          -0.99              -0.81     -0.74
## Happiness.Score          1.00              0.81      0.75
## Whisker.high             1.00              0.81      0.75
## Whisker.low              1.00              0.81      0.75
## Economy..GDP.per.Capita.  0.81              1.00      0.70
## Family                   0.75              0.70      1.00
## Health..Life.Expectancy.  0.79              0.84      0.61
## Freedom                  0.57              0.33      0.42
## Generosity               0.14             -0.06      0.04
## Trust..Government.Corrup. 0.44              0.33      0.24
## Dystopia.Residual        0.51              0.06      0.09
```

## beer_servings	0.51	0.55	0.54
## spirit_servings	0.29	0.35	0.40
## wine_servings	0.50	0.55	0.41
## total_litres_of_pure_alcohol	0.45	0.52	0.52
## population_share	-0.04	-0.01	-0.09
## tv_audience_share	0.14	0.14	0.13
## gdp_weighted_share	0.30	0.33	0.25
##	Health..Life.Expectancy.	Freedom	Generosity
## Happiness.Rank	-0.78	-0.54	-0.12
## Happiness.Score	0.79	0.56	0.14
## Whisker.high	0.78	0.56	0.14
## Whisker.low	0.79	0.57	0.14
## Economy..GDP.per.Capita.	0.84	0.33	-0.06
## Family	0.61	0.42	0.04
## Health..Life.Expectancy.	1.00	0.35	0.03
## Freedom	0.35	1.00	0.32
## Generosity	0.03	0.32	1.00
## Trust..Government.Corruption.	0.26	0.49	0.29
## Dystopia.Residual	0.11	0.12	-0.10
## beer_servings	0.46	0.22	-0.16
## spirit_servings	0.41	-0.03	-0.09
## wine_servings	0.55	0.20	0.00
## total_litres_of_pure_alcohol	0.46	0.14	-0.11
## population_share	0.02	0.05	-0.07
## tv_audience_share	0.18	0.08	-0.07
## gdp_weighted_share	0.36	0.12	-0.07
##	Trust..Government.Corruption.		
## Happiness.Rank	-0.42		
## Happiness.Score	0.44		
## Whisker.high	0.44		
## Whisker.low	0.44		
## Economy..GDP.per.Capita.	0.33		
## Family	0.24		
## Health..Life.Expectancy.	0.26		
## Freedom	0.49		
## Generosity	0.29		
## Trust..Government.Corruption.	1.00		
## Dystopia.Residual	0.04		
## beer_servings	0.05		
## spirit_servings	-0.12		
## wine_servings	0.24		
## total_litres_of_pure_alcohol	0.07		
## population_share	-0.10		
## tv_audience_share	-0.12		
## gdp_weighted_share	0.05		
##	Dystopia.Residual	beer_servings	
## Happiness.Rank	-0.52	-0.51	
## Happiness.Score	0.51	0.51	
## Whisker.high	0.51	0.51	
## Whisker.low	0.51	0.51	
## Economy..GDP.per.Capita.	0.06	0.55	
## Family	0.09	0.54	
## Health..Life.Expectancy.	0.11	0.46	
## Freedom	0.12	0.22	

## Generosity	-0.10	-0.16
## Trust..Government.Corruption.	0.04	0.05
## Dystopia.Residual	1.00	0.14
## beer_servings	0.14	1.00
## spirit_servings	0.01	0.46
## wine_servings	0.08	0.55
## total_litres_of_pure_alcohol	0.06	0.83
## population_share	-0.02	-0.10
## tv_audience_share	0.05	0.06
## gdp_weighted_share	0.07	0.20
##	spirit_servings	wine_servings
## Happiness.Rank	-0.31	-0.49
## Happiness.Score	0.29	0.50
## Whisker.high	0.29	0.49
## Whisker.low	0.29	0.50
## Economy..GDP.per.Capita.	0.35	0.55
## Family	0.40	0.41
## Health..Life.Expectancy.	0.41	0.55
## Freedom	-0.03	0.20
## Generosity	-0.09	0.00
## Trust..Government.Corruption.	-0.12	0.24
## Dystopia.Residual	0.01	0.08
## beer_servings	0.46	0.55
## spirit_servings	1.00	0.24
## wine_servings	0.24	1.00
## total_litres_of_pure_alcohol	0.63	0.68
## population_share	0.09	-0.11
## tv_audience_share	0.16	-0.02
## gdp_weighted_share	0.21	0.20
##	total_litres_of_pure_alcohol	
## Happiness.Rank		-0.46
## Happiness.Score		0.45
## Whisker.high		0.45
## Whisker.low		0.45
## Economy..GDP.per.Capita.		0.52
## Family		0.52
## Health..Life.Expectancy.		0.46
## Freedom		0.14
## Generosity		-0.11
## Trust..Government.Corruption.		0.07
## Dystopia.Residual		0.06
## beer_servings		0.83
## spirit_servings		0.63
## wine_servings		0.68
## total_litres_of_pure_alcohol		1.00
## population_share		-0.06
## tv_audience_share		0.08
## gdp_weighted_share		0.26
##	population_share	tv_audience_share
## Happiness.Rank	0.04	-0.15
## Happiness.Score	-0.04	0.14
## Whisker.high	-0.05	0.13
## Whisker.low	-0.04	0.14
## Economy..GDP.per.Capita.	-0.01	0.14

## Family	-0.09	0.13
## Health..Life.Expectancy.	0.02	0.18
## Freedom	0.05	0.08
## Generosity	-0.07	-0.07
## Trust..Government.Corruption.	-0.10	-0.12
## Dystopia.Residual	-0.02	0.05
## beer_servings	-0.10	0.06
## spirit_servings	0.09	0.16
## wine_servings	-0.11	-0.02
## total_litres_of_pure_alcohol	-0.06	0.08
## population_share	1.00	0.72
## tv_audience_share	0.72	1.00
## gdp_weighted_share	0.42	0.80
##	gdp_weighted_share	
## Happiness.Rank	-0.32	
## Happiness.Score	0.30	
## Whisker.high	0.29	
## Whisker.low	0.30	
## Economy..GDP.per.Capita.	0.33	
## Family	0.25	
## Health..Life.Expectancy.	0.36	
## Freedom	0.12	
## Generosity	-0.07	
## Trust..Government.Corruption.	0.05	
## Dystopia.Residual	0.07	
## beer_servings	0.20	
## spirit_servings	0.21	
## wine_servings	0.20	
## total_litres_of_pure_alcohol	0.26	
## population_share	0.42	
## tv_audience_share	0.80	
## gdp_weighted_share	1.00	
##		
## n= 137		
##		
##		
## P		
##	Happiness.Rank	Happiness.Score Whisker.high
## Happiness.Rank	0.0000	0.0000
## Happiness.Score	0.0000	0.0000
## Whisker.high	0.0000	0.0000
## Whisker.low	0.0000	0.0000
## Economy..GDP.per.Capita.	0.0000	0.0000
## Family	0.0000	0.0000
## Health..Life.Expectancy.	0.0000	0.0000
## Freedom	0.0000	0.0000
## Generosity	0.1628	0.0932
## Trust..Government.Corruption.	0.0000	0.0000
## Dystopia.Residual	0.0000	0.0000
## beer_servings	0.0000	0.0000
## spirit_servings	0.0003	0.0006
## wine_servings	0.0000	0.0000
## total_litres_of_pure_alcohol	0.0000	0.0000
## population_share	0.6081	0.6041
		0.5562

## tv_audience_share	0.0802	0.1128	0.1248
## gdp_weighted_share	0.0002	0.0004	0.0005
##	Whisker.low	Economy..GDP.per.Capita.	Family
## Happiness.Rank	0.0000	0.0000	0.0000
## Happiness.Score	0.0000	0.0000	0.0000
## Whisker.high	0.0000	0.0000	0.0000
## Whisker.low		0.0000	0.0000
## Economy..GDP.per.Capita.	0.0000		0.0000
## Family	0.0000	0.0000	
## Health..Life.Expectancy.	0.0000	0.0000	0.0000
## Freedom	0.0000	0.0000	0.0000
## Generosity	0.0944	0.5223	0.6216
## Trust..Government.Corruption.	0.0000	0.0000	0.0051
## Dystopia.Residual	0.0000	0.4774	0.3217
## beer_servings	0.0000	0.0000	0.0000
## spirit_servings	0.0005	0.0000	0.0000
## wine_servings	0.0000	0.0000	0.0000
## total_litres_of_pure_alcohol	0.0000	0.0000	0.0000
## population_share	0.6528	0.8698	0.2708
## tv_audience_share	0.1023	0.0945	0.1374
## gdp_weighted_share	0.0003	0.0000	0.0028
##	Health..Life.Expectancy.	Freedom	Generosity
## Happiness.Rank	0.0000	0.0000	0.1628
## Happiness.Score	0.0000	0.0000	0.0932
## Whisker.high	0.0000	0.0000	0.0923
## Whisker.low	0.0000	0.0000	0.0944
## Economy..GDP.per.Capita.	0.0000	0.0000	0.5223
## Family	0.0000	0.0000	0.6216
## Health..Life.Expectancy.		0.0000	0.6883
## Freedom	0.0000		0.0001
## Generosity	0.6883	0.0001	
## Trust..Government.Corruption.	0.0019	0.0000	0.0007
## Dystopia.Residual	0.2219	0.1609	0.2636
## beer_servings	0.0000	0.0109	0.0680
## spirit_servings	0.0000	0.6863	0.2761
## wine_servings	0.0000	0.0169	0.9729
## total_litres_of_pure_alcohol	0.0000	0.1094	0.1910
## population_share	0.7883	0.5465	0.3910
## tv_audience_share	0.0401	0.3326	0.3963
## gdp_weighted_share	0.0000	0.1514	0.4102
##	Trust..Government.Corruption.		
## Happiness.Rank	0.0000		
## Happiness.Score	0.0000		
## Whisker.high	0.0000		
## Whisker.low	0.0000		
## Economy..GDP.per.Capita.	0.0000		
## Family	0.0051		
## Health..Life.Expectancy.	0.0019		
## Freedom	0.0000		
## Generosity	0.0007		
## Trust..Government.Corruption.			
## Dystopia.Residual	0.6504		
## beer_servings	0.5262		
## spirit_servings	0.1693		

```

## wine_servings          0.0048
## total_litres_of_pure_alcohol 0.4081
## population_share       0.2271
## tv_audience_share     0.1706
## gdp_weighted_share     0.5508
## Dystopia.Residual      beer_servings
## Happiness.Rank         0.0000          0.0000
## Happiness.Score        0.0000          0.0000
## Whisker.high           0.0000          0.0000
## Whisker.low            0.0000          0.0000
## Economy..GDP.per.Capita. 0.4774          0.0000
## Family                 0.3217          0.0000
## Health..Life.Expectancy. 0.2219          0.0000
## Freedom                0.1609          0.0109
## Generosity              0.2636          0.0680
## Trust..Government.Corruption. 0.6504          0.5262
## Dystopia.Residual      0.0944
## beer_servings          0.0944
## spirit_servings         0.9490          0.0000
## wine_servings          0.3651          0.0000
## total_litres_of_pure_alcohol 0.4702          0.0000
## population_share       0.8149          0.2536
## tv_audience_share     0.5273          0.4670
## gdp_weighted_share     0.4403          0.0191
## spirit_servings         wine_servings
## Happiness.Rank         0.0003          0.0000
## Happiness.Score        0.0006          0.0000
## Whisker.high           0.0007          0.0000
## Whisker.low            0.0005          0.0000
## Economy..GDP.per.Capita. 0.0000          0.0000
## Family                 0.0000          0.0000
## Health..Life.Expectancy. 0.0000          0.0000
## Freedom                0.6863          0.0169
## Generosity              0.2761          0.9729
## Trust..Government.Corruption. 0.1693          0.0048
## Dystopia.Residual      0.9490          0.3651
## beer_servings          0.0000          0.0000
## spirit_servings         0.0052
## wine_servings          0.0052
## total_litres_of_pure_alcohol 0.0000          0.0000
## population_share       0.2920          0.2189
## tv_audience_share     0.0680          0.8286
## gdp_weighted_share     0.0121          0.0203
## total_litres_of_pure_alcohol
## Happiness.Rank         0.0000
## Happiness.Score        0.0000
## Whisker.high           0.0000
## Whisker.low            0.0000
## Economy..GDP.per.Capita. 0.0000
## Family                 0.0000
## Health..Life.Expectancy. 0.0000
## Freedom                0.1094
## Generosity              0.1910
## Trust..Government.Corruption. 0.4081

```

## Dystopia.Residual	0.4702	
## beer_servings	0.0000	
## spirit_servings	0.0000	
## wine_servings	0.0000	
## total_litres_of_pure_alcohol		
## population_share	0.4623	
## tv_audience_share	0.3515	
## gdp_weighted_share	0.0026	
##	population_share	tv_audience_share
## Happiness.Rank	0.6081	0.0802
## Happiness.Score	0.6041	0.1128
## Whisker.high	0.5562	0.1248
## Whisker.low	0.6528	0.1023
## Economy..GDP.per.Capita.	0.8698	0.0945
## Family	0.2708	0.1374
## Health..Life.Expectancy.	0.7883	0.0401
## Freedom	0.5465	0.3326
## Generosity	0.3910	0.3963
## Trust..Government.Corruption.	0.2271	0.1706
## Dystopia.Residual	0.8149	0.5273
## beer_servings	0.2536	0.4670
## spirit_servings	0.2920	0.0680
## wine_servings	0.2189	0.8286
## total_litres_of_pure_alcohol	0.4623	0.3515
## population_share		0.0000
## tv_audience_share	0.0000	
## gdp_weighted_share	0.0000	0.0000
##	gdp_weighted_share	
## Happiness.Rank	0.0002	
## Happiness.Score	0.0004	
## Whisker.high	0.0005	
## Whisker.low	0.0003	
## Economy..GDP.per.Capita.	0.0000	
## Family	0.0028	
## Health..Life.Expectancy.	0.0000	
## Freedom	0.1514	
## Generosity	0.4102	
## Trust..Government.Corruption.	0.5508	
## Dystopia.Residual	0.4403	
## beer_servings	0.0191	
## spirit_servings	0.0121	
## wine_servings	0.0203	
## total_litres_of_pure_alcohol	0.0026	
## population_share	0.0000	
## tv_audience_share	0.0000	
## gdp_weighted_share		



It's important to note that this correlation matrix does not include the Trump data as there are too many missing values, and the country variable has been removed for the time being. Now let's focus on making the matrix more digestible by flattening the values. For the sake of viewing, I've elected to only show the top 25 observations.

```
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor = (cormat)[ut],
    p = pmat[ut]
  )
}
library(Hmisc)
flatdata<- flattenCorrMatrix(corr1$r, corr1$p)
head(flatdata, n=25)
```

row	column	cor	p
Happiness.Rank	Happiness.Score	-0.9925736	0
Happiness.Rank	Whisker.high	-0.9927633	0
Happiness.Score	Whisker.high	0.9995732	0
Happiness.Rank	Whisker.low	-0.9915725	0
Happiness.Score	Whisker.low	0.9995941	0
Whisker.high	Whisker.low	0.9983350	0
Happiness.Rank	Economy..GDP.per.Capita.	-0.8057458	0
Happiness.Score	Economy..GDP.per.Capita.	0.8066217	0
Whisker.high	Economy..GDP.per.Capita.	0.8064064	0
Whisker.low	Economy..GDP.per.Capita.	0.8061685	0
Happiness.Rank	Family	-0.7364378	0
Happiness.Score	Family	0.7518377	0
Whisker.high	Family	0.7509724	0
Whisker.low	Family	0.7520635	0
Economy..GDP.per.Capita.	Family	0.6977798	0
Happiness.Rank	Health..Life.Expectancy.	-0.7848782	0
Happiness.Score	Health..Life.Expectancy.	0.7852965	0
Whisker.high	Health..Life.Expectancy.	0.7807859	0
Whisker.low	Health..Life.Expectancy.	0.7890496	0
Economy..GDP.per.Capita.	Health..Life.Expectancy.	0.8432045	0
Family	Health..Life.Expectancy.	0.6117689	0
Happiness.Rank	Freedom	-0.5446922	0
Happiness.Score	Freedom	0.5649702	0
Whisker.high	Freedom	0.5638798	0
Whisker.low	Freedom	0.5655690	0

Based off of a basic correlation matrix there is a plausible relationship between the amount of alcohol consumed and the various metrics of happiness. In particular, total beer and wine consumption by a country seems to score well with happiness index and the related variables such as GDP per capita, and family happiness. Unfortunately we see now relationship between alcohol consumption of any sort and its ability to affect soccer fandom in a country. This shouldn't be surprising because soccer is a universal sport that penetrates cultures that largely abstain from drinking. What could be interesting is conducted an analysis by looking at different country regions, but let's get back to that later. First let's see what we can dig up about beer and wine consumption and their effect upon a country's happiness score.

Let's see compare the beer and wine consumption for the top 10 happiest and least happy countries, starting with beer. First, let's sort the data.

```
sortHappy <- TrumpMerge[order(TrumpMerge$Happiness.Rank),]  
sortunhappy <- TrumpMerge[order(-TrumpMerge$Happiness.Rank),]
```

Now let take the top 5 of the happiest and unhappiest countries

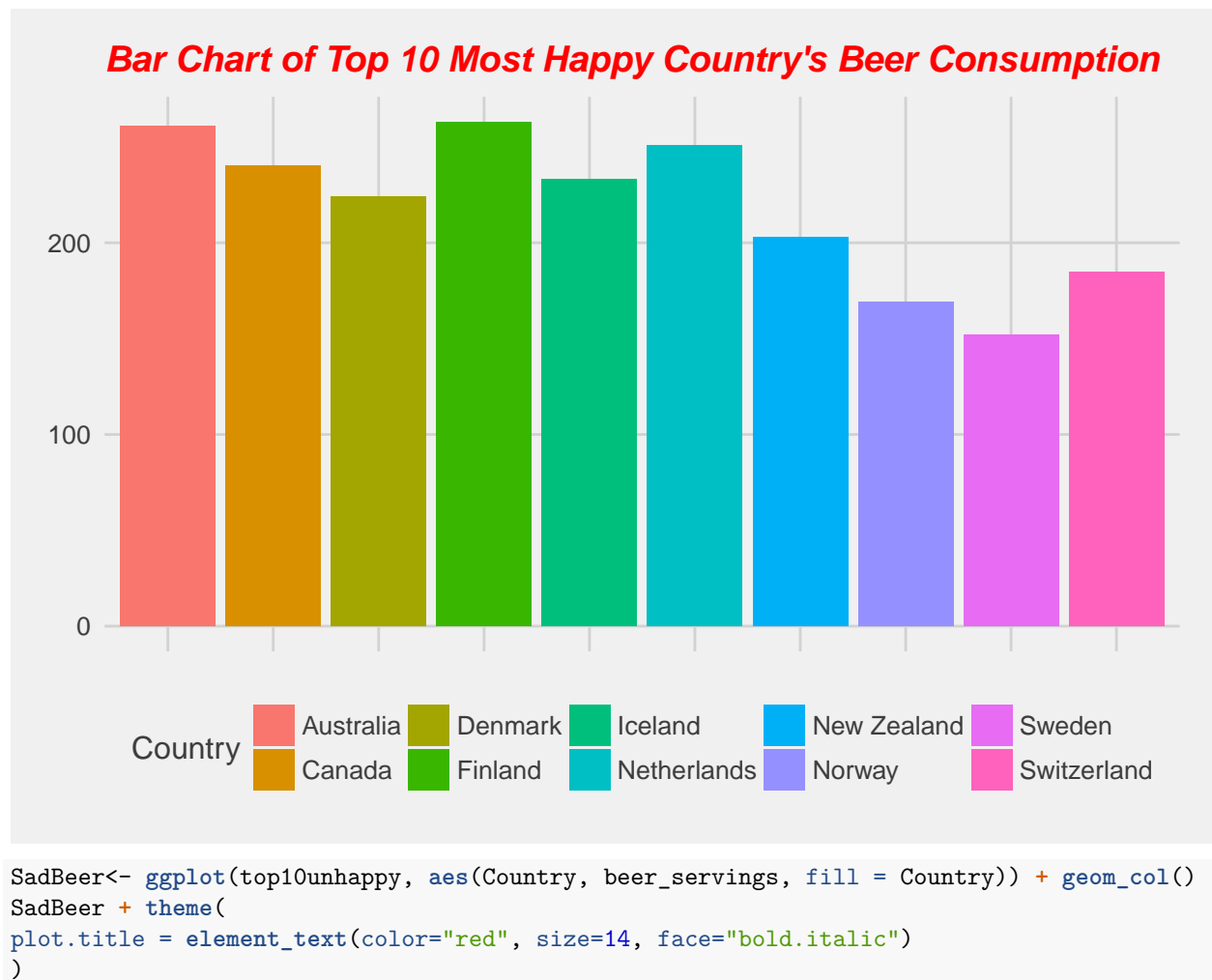
```
top10happy <- sortHappy[1:10,]  
top10unhappy <- sortunhappy[1:10,]
```

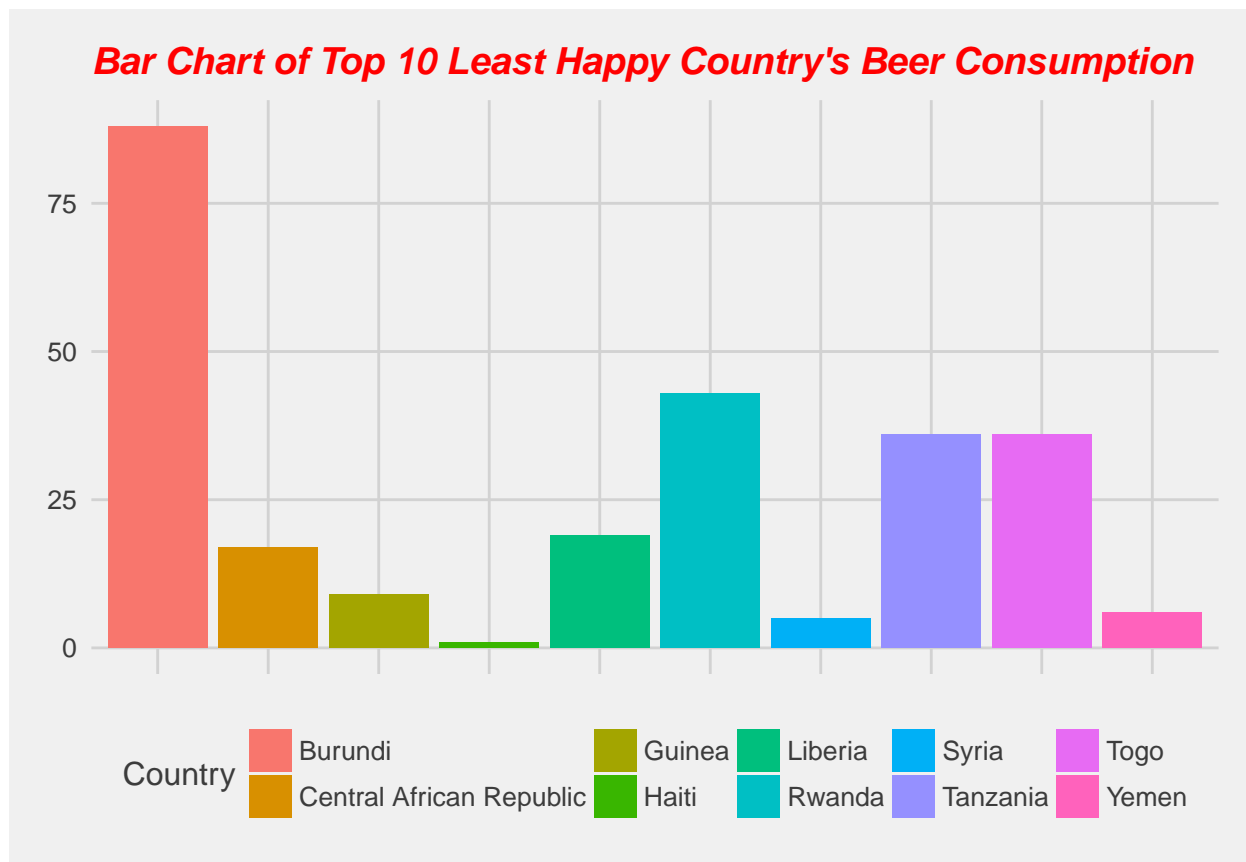
Let's take a look at their beer consumption (cans of beer per year per capita), using my favorite theme!

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 3.3.3
```

```
HappyBeer <- ggplot(top10happy, aes(Country, beer_servings, fill = Country)) + geom_col() + theme_fiv  
HappyBeer + theme(  
plot.title = element_text(color="red", size=14, face="bold.italic")  
)
```





Certainly seems like there is a large disparity of happiness between countries that drink and countries that do not drink a large amount of beer. However that could also be due to gdp. What if we do the same with countries more from the middle of the pack? Let's find out.

A good strategy could be to find where the first and third quartiles are in the happiness scale and fish 10 countries from there.

First let's check the IQR for the Happiness Index

```
summary(TrumpMerge$Happiness.Score)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
2.693	4.497	5.273	5.341	6.098	7.537	1

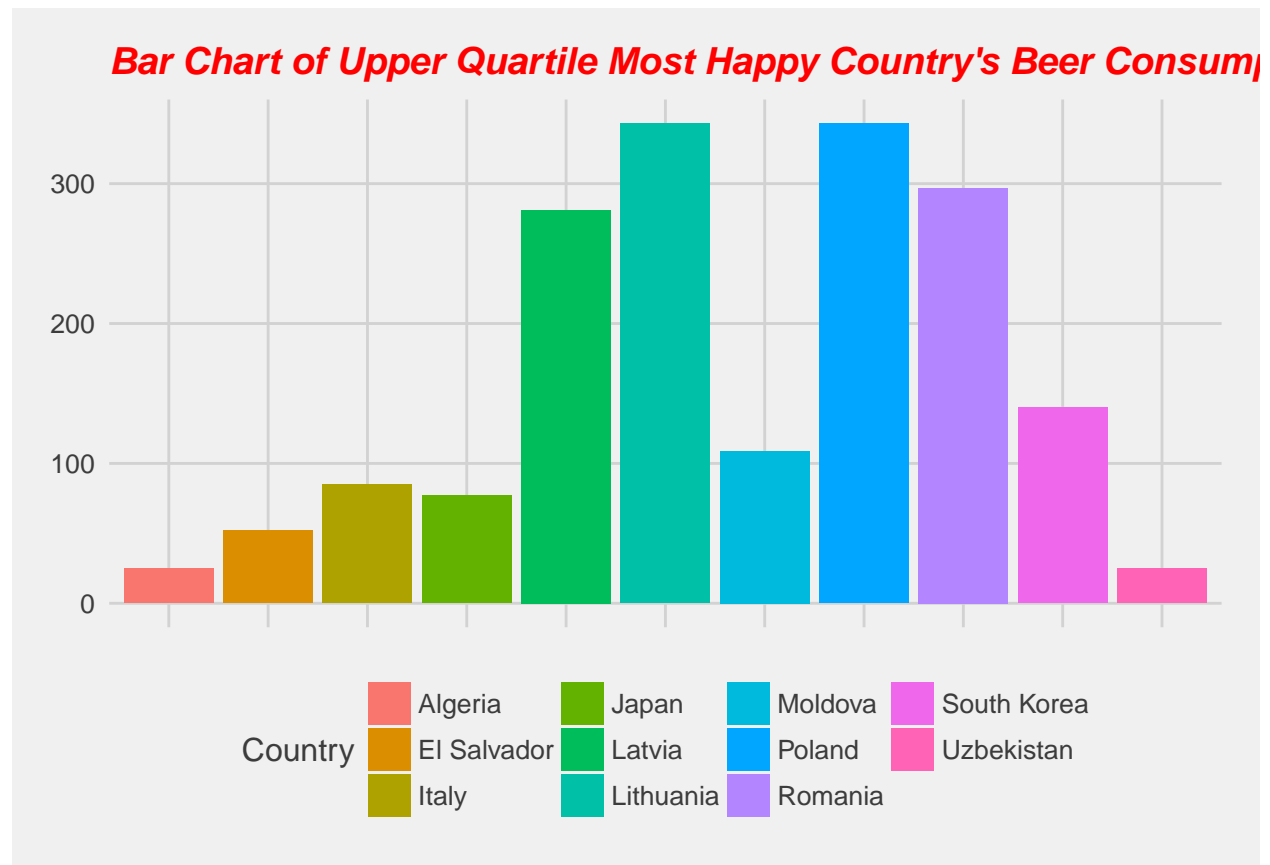
With theses quartiles we're looking at happiness ranks of 40-50 for the 3rd quartile and 100-110 for the 1st, so let's see if the beer consumption is markedly higher for the happier countries.

Let's go ahead and subset the data we wish to have for this bit of data visualization

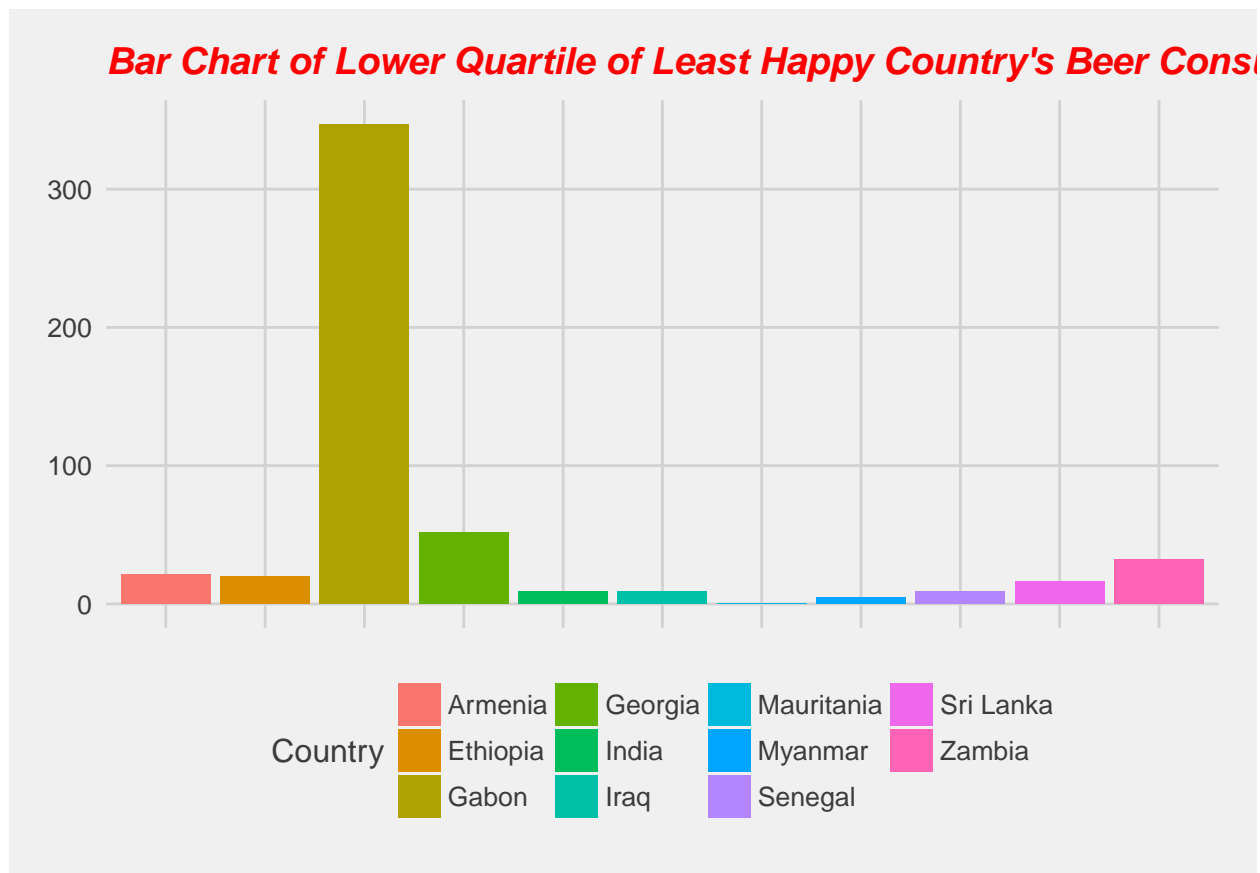
```
Q3Happy <- sortHappy[40:50,]
Q1UnHappy <- sortHappy[100:110,]
```

Now let's observe the bar graphs side by side

```
Q3Beer <- ggplot(Q3Happy, aes(Country, beer_servings, fill = Country)) + geom_col() + theme_fivethirtysix
Q3Beer + theme(
  plot.title = element_text(color="red", size=14, face="bold.italic")
)
```



```
Q1Beer<- ggplot(Q1UnHappy, aes(Country, beer_servings, fill = Country)) + geom_col() + scale_color_fivethirtysix
Q1Beer + theme(
  plot.title = element_text(color="red", size=14, face="bold.italic")
)
```

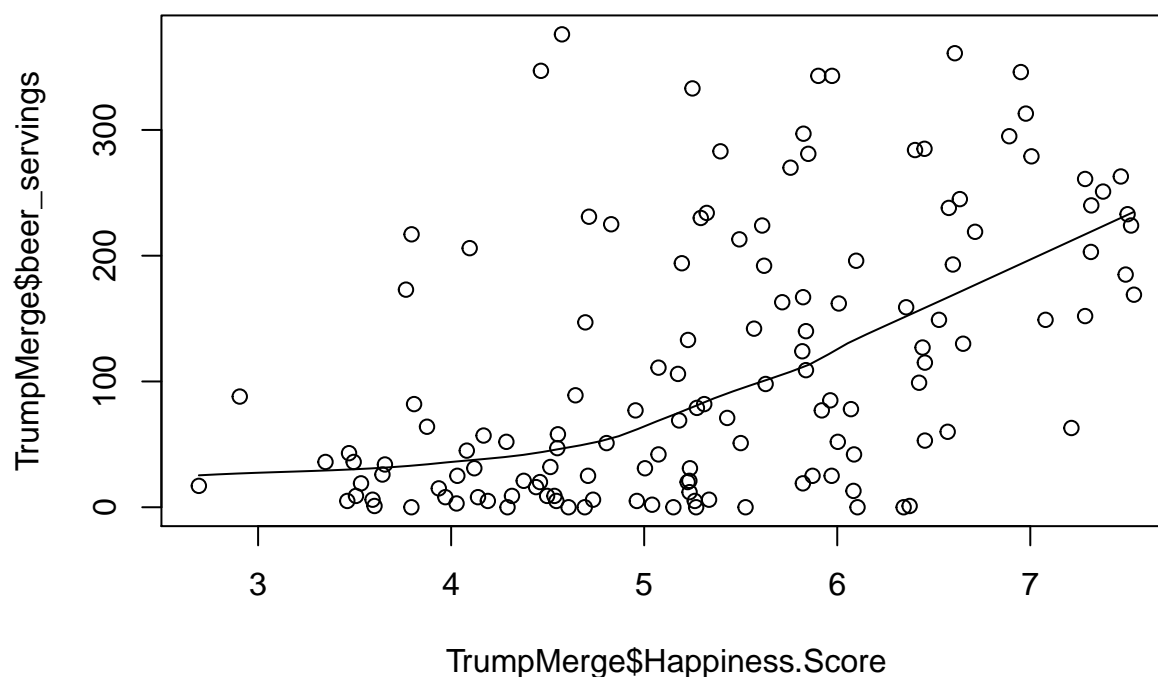


Overall, we still see that the happier countries consume more beer than the unhappy countries, but now we're seeing countries that go against that overall relationship. What could be best moving forward and conducting a crude regression analysis to observe if statistically speaking, we see a linear relationship between annual per capita beer + wine consumption versus a country's happiness score. Afterwards we'll start digging into the plausibility of a relationship between these co-variables and Trump's approval.

Let's first visualize the relationship through a scatter plot diagram of Happiness vs Beer (Annual cans of beer consumed per capita)

```
scatter.smooth(x=TrumpMerge$Happiness.Score, y=TrumpMerge$beer_servings, main="Happiness Score vs Annual Beer Consumption")
```

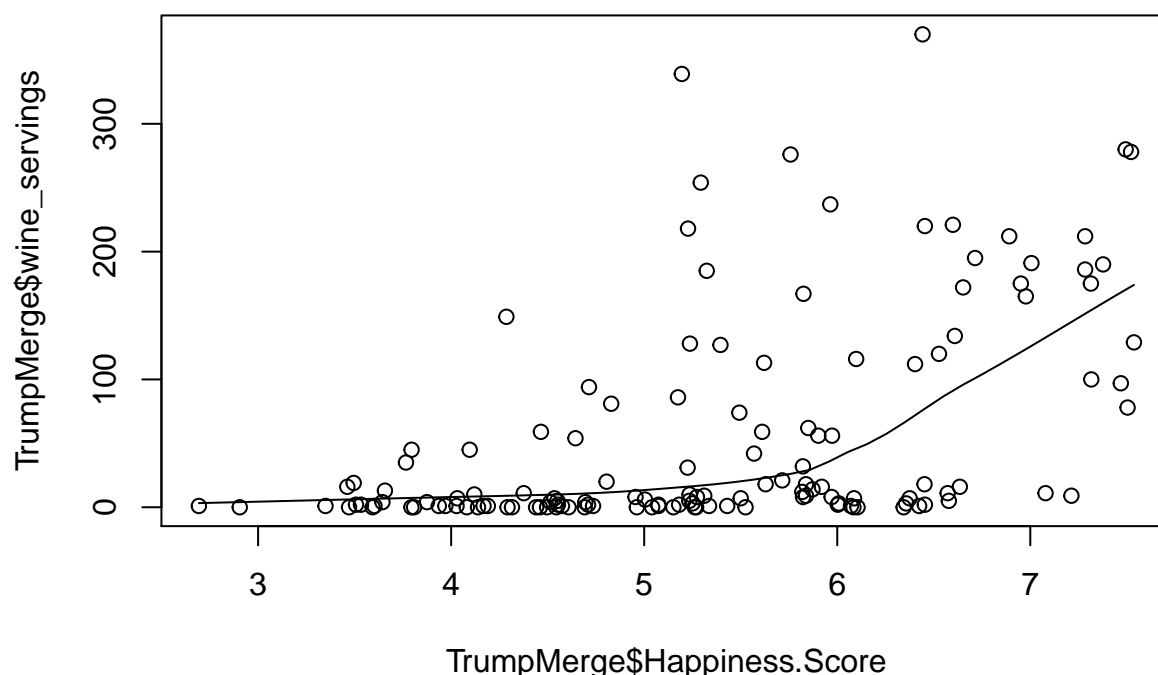
## Happiness Score vs Annual Cans of Beer per Capita



Now let's do the same with Happiness vs Wine consumption (annual glasses of wine consumed per capita)

```
scatter.smooth(x=TrumpMerge$Happiness.Score, y=TrumpMerge$wine_servings, main="Happiness.Score Vs Annual
```

## Happiness.Score Vs Annual glasses of wine per capita



From what it appears there is the plausibility of a linear relationship, let's run the actual regression.

```
linearbeer <- lm(Happiness.Score ~ beer_servings, data = TrumpMerge)
linearwine <- lm(Happiness.Score ~ wine_servings, data = TrumpMerge)
summary(linearbeer)
```

```
##
## Call:
## lm(formula = Happiness.Score ~ beer_servings, data = TrumpMerge)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.29928 -0.68846  0.06917  0.76413  2.14555
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.7226462  0.1239787  38.092 < 2e-16 ***
## beer_servings  0.0054731  0.0007981   6.858 2.28e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9954 on 135 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.2584, Adjusted R-squared:  0.2529
## F-statistic: 47.03 on 1 and 135 DF, p-value: 2.283e-10
```



```
summary(linearwine)
```

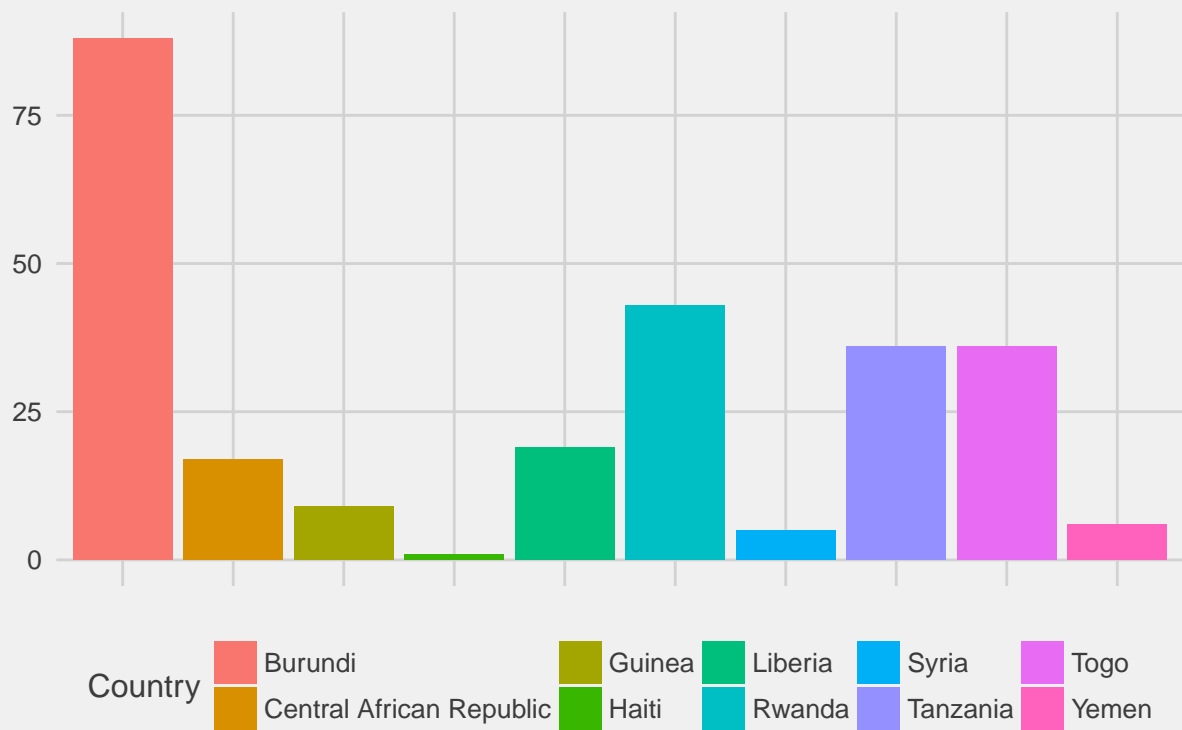
```
##
## Call:
## lm(formula = Happiness.Score ~ wine_servings, data = TrumpMerge)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.27468 -0.79968  0.08013  0.74931  2.19087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.960872   0.102843  48.238 < 2e-16 ***
## wine_servings 0.006807   0.001019   6.678 5.78e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.002 on 135 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.2483, Adjusted R-squared:  0.2427
## F-statistic: 44.59 on 1 and 135 DF,  p-value: 5.781e-10
```

Very interesting, both are significant. According to the beer model, we see that for every can of beer consumed per capita during a year, it increases the happiness score of a country 0.005 points. If we multiply that by 100, we see that every 100 additional beers consumed per capital during a year increases a country's happiness score by 0.5. Considering that the range is between 3 and 7, this is pretty significant. It would also appear that wine is also more effective at influencing the happiness score, with a 0.6 gain per 100 glasses consumed annually per capita.

This should make sense though, a large component of a country's happiness score is its GDP per capita and this influences the ability of a country's citizens to purchase alcohol, especially affordable forms such as beer and wine. It's also important to consider the fact that this data is taken from legal sales of alcohol. Countries with less available monetary capital rely more on illegal forms of alcohol for consumption (WHO 2011). The "Global status report on alcohol and health" produced by the WHO also goes into greater detail in describing this phenomenon. Noticed earlier how we saw that most of the countries that were at the lower tiers of happiness index had such low alcohol consumption were developing nations:

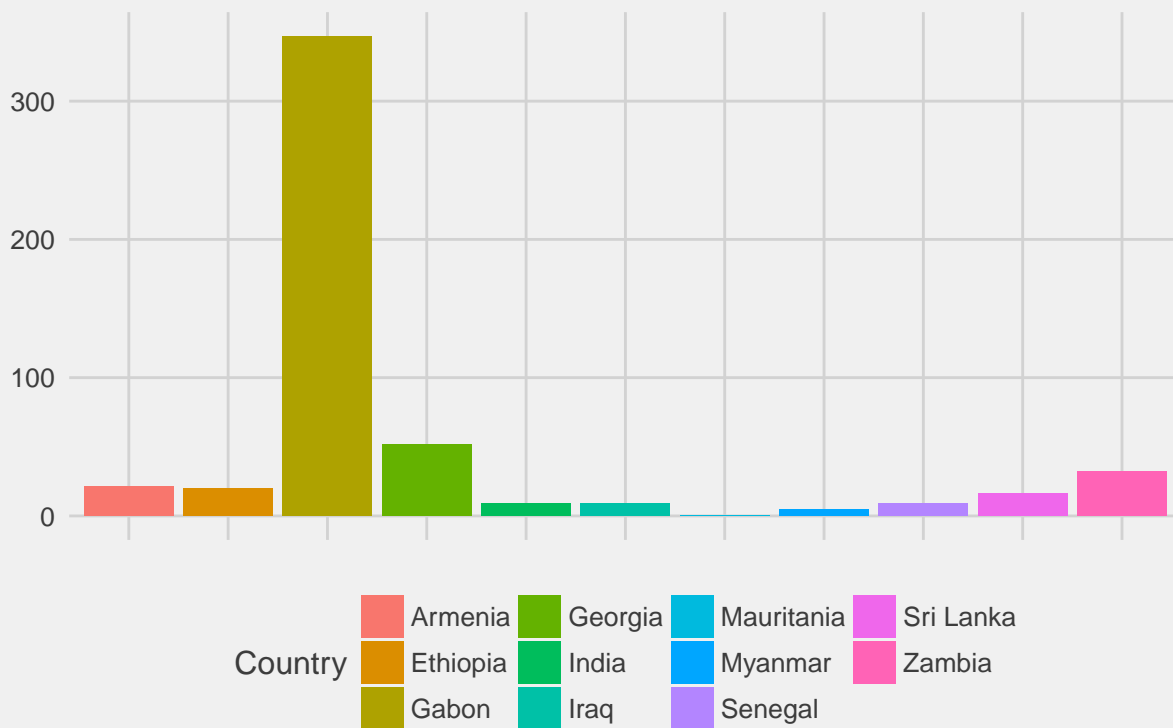
```
SadBeer + theme(
plot.title = element_text(color="red", size=14, face="bold.italic")
)
```

***Bar Chart of Top 10 Least Happy Country's Beer Consumption***



```
Q1Beer + theme(  
  plot.title = element_text(color="red", size=14, face="bold.italic")  
)
```

**Bar Chart of Lower Quartile of Least Happy Country's Beer Cons**



According to the report, countries in Africa and South East Asia regions typically have citizens make their own form of alcohol that could be distributed in an informal market setting which would reduce the official consumption numbers used in this data set. This dataset also, of course, favors countries with stronger surveillance capabilities. Countries above such as Haiti have used their resources to surveillance of communicable diseases and vaccination over point of sales data (CDC 2013).

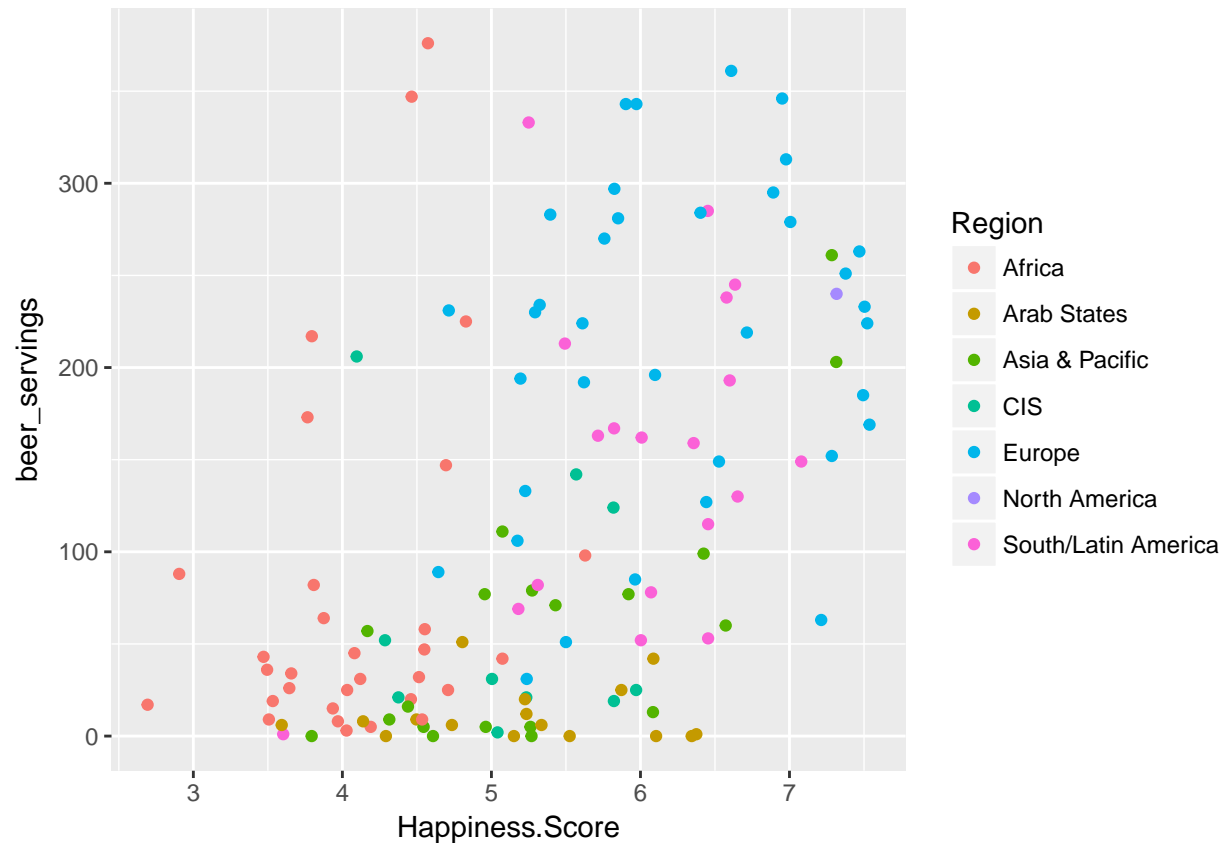
Maybe if we divide the global regions up we could see the differences in both happiness and beer/wine consumption that could explain this behavior. Maybe one region is accounting for more of this effect than others. Let's explore this.

First we need to assign the countries to a region, for this we will need to merge a list of countries and their respective region. For the purpose of this analysis we will be using a list provided by the International Telecommunications Union (ITU 2018).

```
final <- merge(TrumpMerge, region, by = "Country")
```

Let's display the happiness and beerconsumption by region through a colored scatterplot to get a better idea where each region falls within the greater distribution of our data

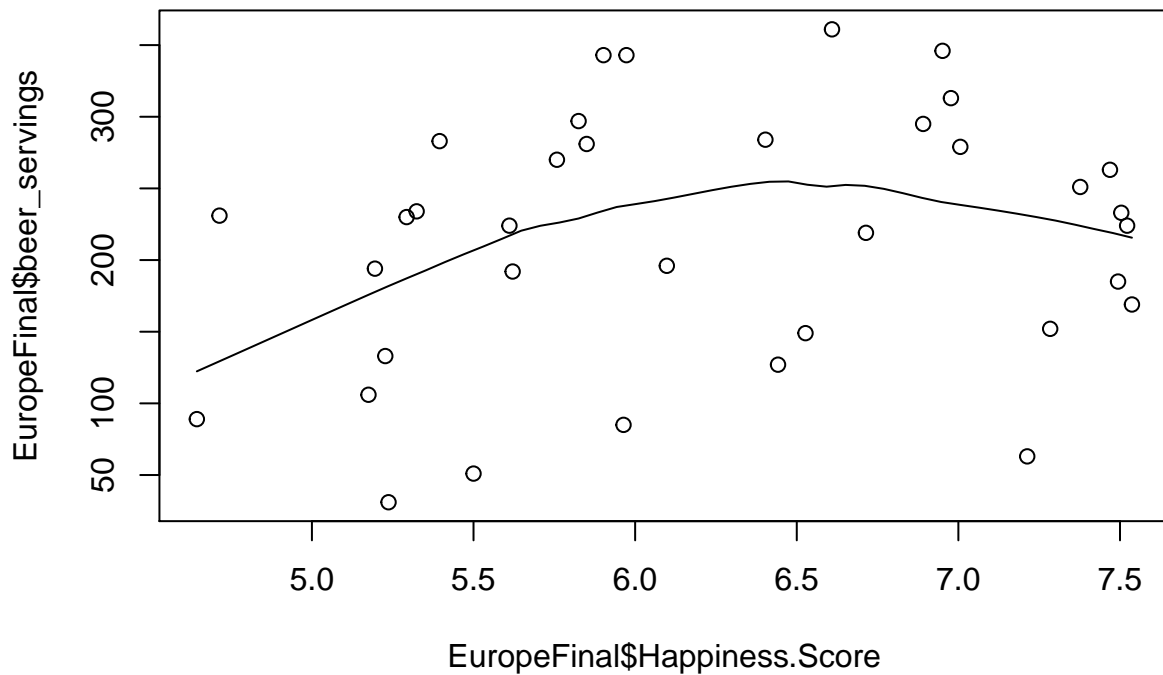
```
qplot(Happiness.Score, beer_servings, colour = Region,
      data = final)
```



Well it seems that the countries that drink beer often and are happier are all typically found in the region of Europe, however, we do see some Latin American countries driving this statistical relationship as well. Maybe we should look at this relationship per region. Let's break them down starting by region.

```
EuropeFinal <-subset(final,Region %in% 'Europe')
scatter.smooth(x=EuropeFinal$Happiness.Score, y=EuropeFinal$beer_servings, main="Happiness vs Beer in Europe")
```

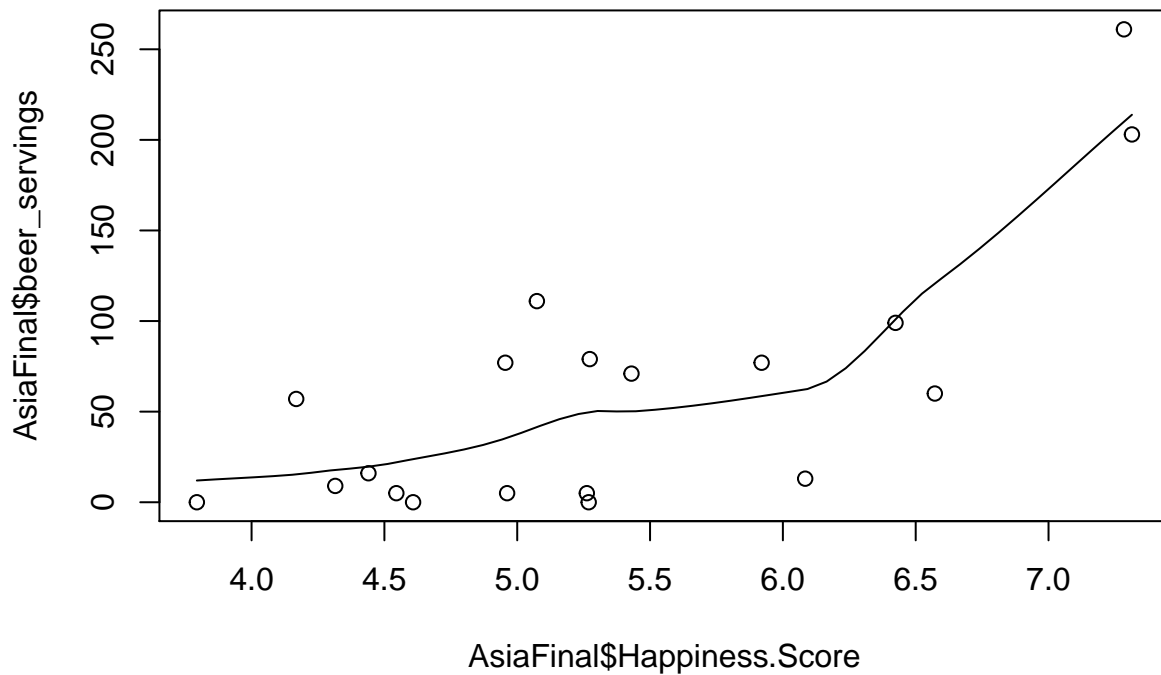
## Happiness vs Beer in Europe



The trend line suggests there is a clear relationship until about 250 beers, then we see diminishing returns. Then again Europe is a country with a healthy mix of wine, spirit, and beer sales, but we see that they are indeed, a large part of the overall trend we saw in the aggregate scatter plot. We see that the happiest countries that had high beer consumption were from Europe, Asia, and North America. Let's look at Asia next.

```
AsiaFinal <-subset(final,Region %in% 'Asia & Pacific')
scatter.smooth(x=AsiaFinal$Happiness.Score, y=AsiaFinal$beer_servings, main="Happiness vs Beer in Asia & Pacific")
```

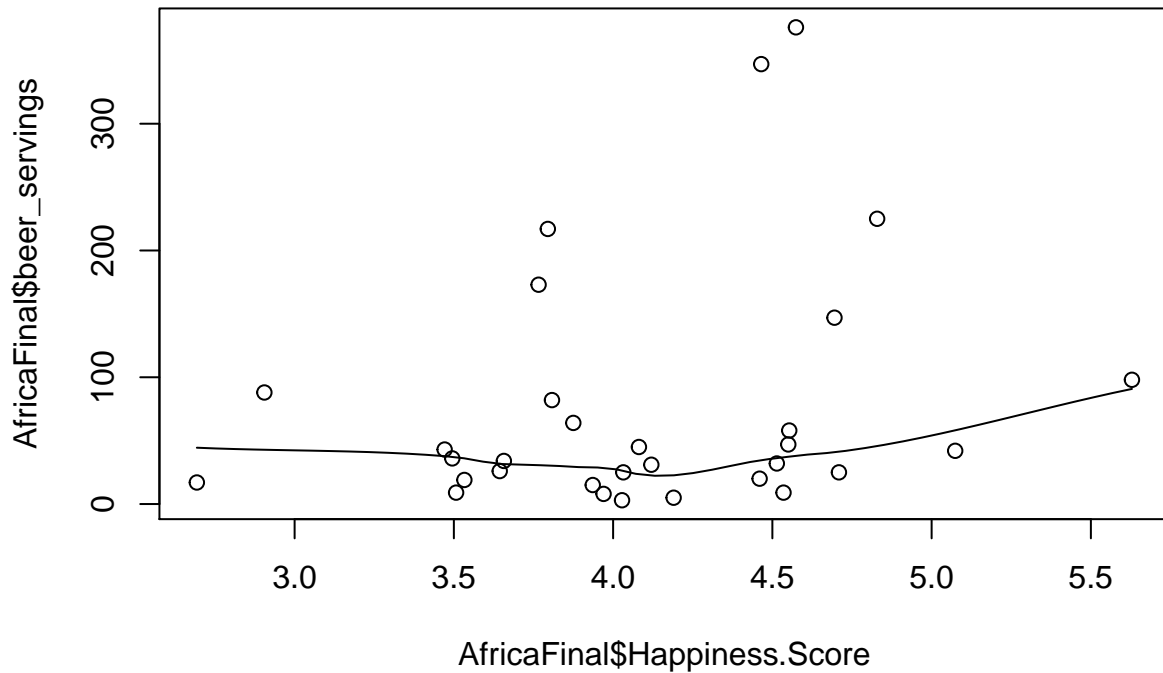
## Happiness vs Beer in Asia & Pacific Region



This trend line also is increasingly positive and we see why these two regions provide a lot of the statistical power behind the relationship between beer and happiness. On the opposite end we see that the least happy countries that drink very little beer are in Africa. Of course, GDP is one of the factors of the happiness index and thus it is not surprising to find Africa near the bottom. Perhaps, this region adds onto the trend as well. Let's observe the trend line.

```
AfricaFinal <-subset(final,Region %in% 'Africa')
scatter.smooth(x=AfricaFinal$Happiness.Score, y=AfricaFinal$beer_servings, main="Happiness vs Beer in A
```

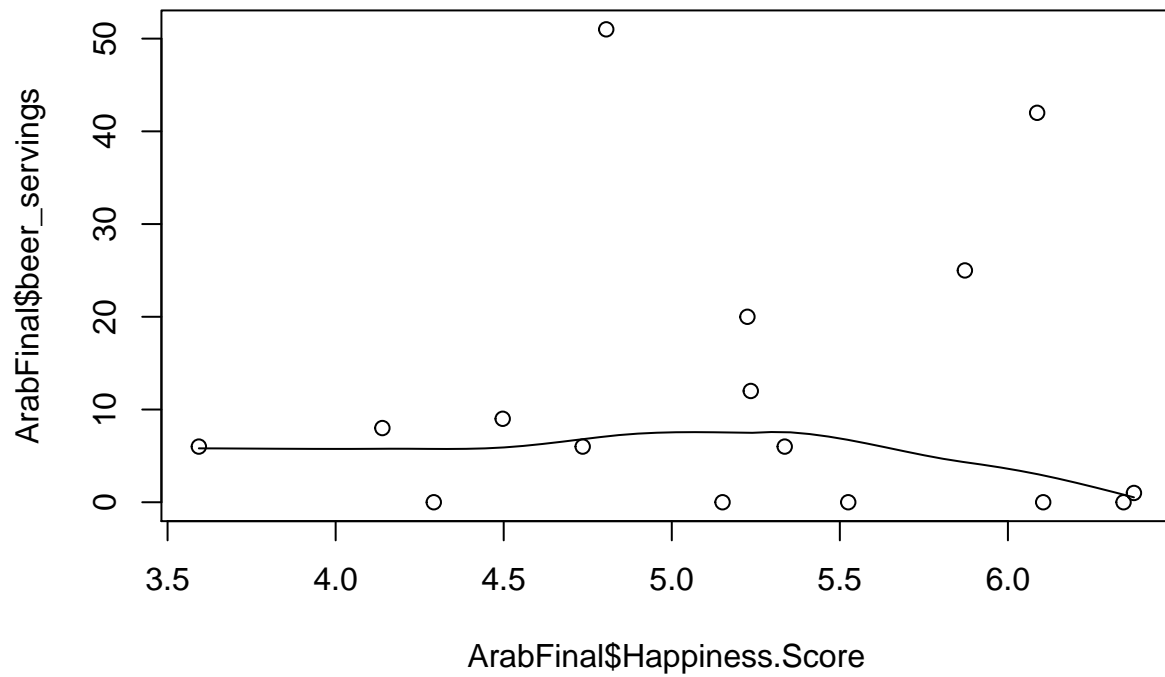
## Happiness vs Beer in Africa Region



As predicted, Africa as a region does not have a significant positive association between Happiness and the total amount of beer consumed. Thus far it would seem that the Europe and Asia are the two biggest drivers of this trend. Let's final out if the final two regions with a sample size larger than 15 (Arab States and Latin/South America)

```
ArabFinal <-subset(final,Region %in% 'Arab States')
LatinFinal <- subset(final, Region %in% 'South/Latin America')
scatter.smooth(x=ArabFinal$Happiness.Score, y=ArabFinal$beer_servings, main="Happiness vs Beer in Arab L
```

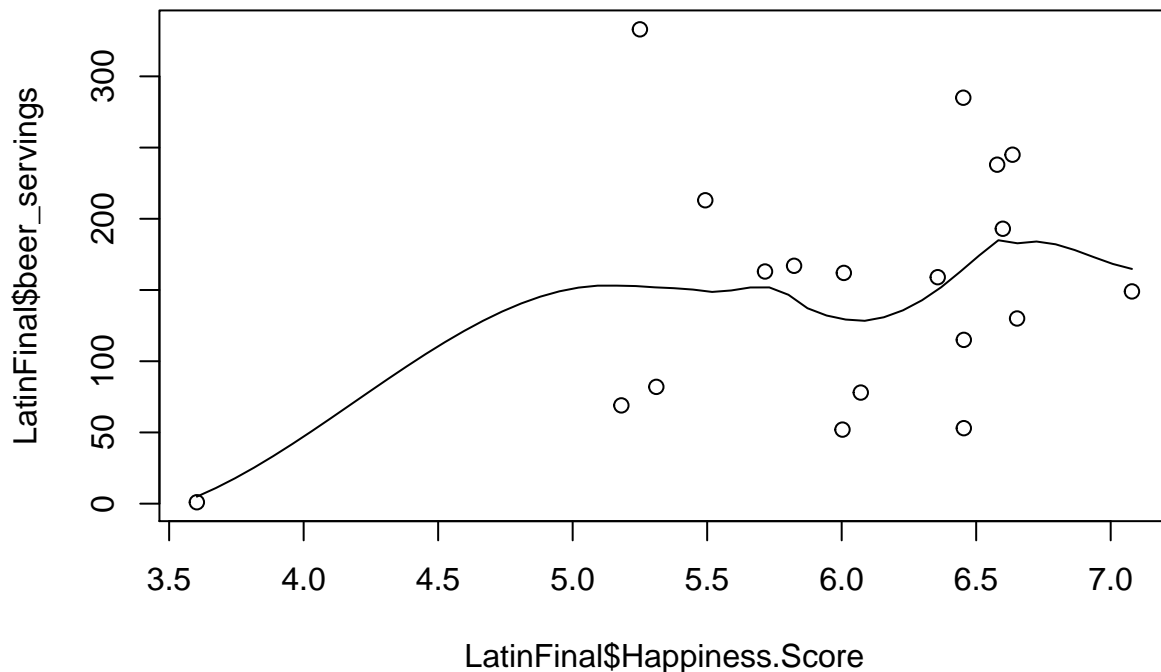
## Happiness vs Beer in Arab Region



```
scatter.smooth(x=LatinFinal$Happiness.Score, y=LatinFinal$beer_servings, main="Happiness vs Beer in Lat
```



## Happiness vs Beer in Latin Region



We see that the Arab Region does not have any trend on the influence of beer consumption upon Happiness and that the Latin region has a small trend line that mostly driven by the presence of a single outlying point.

We've now learned that the trend between happiness and beer consumption is strongest in Europe and Asia, weak in Latin America, and weakest in Africa and the Arab States. This is good to know because now when we try to compare these two variables to Trump's approval, we can have an idea if this sort of regionality still exists.

Let's isolate only complete observations that have approval ratings for Trump, to do this a bit of data cleaning is necessary.

```
final$X3 <- NULL
final2 <- final[complete.cases(final), ]
```

With our dataset in hand let's compare the net approval rating of the 34 remaining countries compared to their overall happiness.

```
TrumpHappy <- lm(net_approval ~ Happiness.Score, data = final2)
summary(TrumpHappy)
```

```
##
## Call:
## lm(formula = net_approval ~ Happiness.Score, data = final2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -36.656 -14.342   2.184  11.980  38.383
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      26.70      18.76   1.423 0.164338
## Happiness.Score  -13.46       3.16  -4.259 0.000168 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.35 on 32 degrees of freedom
## Multiple R-squared:  0.3618, Adjusted R-squared:  0.3418
## F-statistic: 18.14 on 1 and 32 DF,  p-value: 0.0001684
```

Wow, well as it turns out, the Happier the country, the less likely it approves of trump. Let's quickly explore if other variables in our dataset is predictive of this, particularly alcohol consumption.

```
library(MASS)
TrumpTest <- lm(net_approval ~ Happiness.Score + beer_servings + wine_servings + spirit_servings + total_litres_of_pure_alcohol)
step <- stepAIC(TrumpTest, direction="both")
```

```
## Start:  AIC=197.61
## net_approval ~ Happiness.Score + beer_servings + wine_servings +
##      spirit_servings + total_litres_of_pure_alcohol
##
##               Df Sum of Sq    RSS    AIC
## - spirit_servings      1      25.79 8012.1 195.72
## - beer_servings       1     105.20 8091.5 196.06
## - total_litres_of_pure_alcohol 1     142.38 8128.7 196.21
## <none>                        7986.3 197.61
## - wine_servings       1    1208.72 9195.0 200.40
## - Happiness.Score     1    1371.00 9357.3 201.00
##
## Step:  AIC=195.72
## net_approval ~ Happiness.Score + beer_servings + wine_servings +
##      total_litres_of_pure_alcohol
##
##               Df Sum of Sq    RSS    AIC
## - beer_servings      1      96.97 8109.1 194.13
## - total_litres_of_pure_alcohol 1     118.15 8130.3 194.22
## <none>                        8012.1 195.72
## + spirit_servings     1      25.79 7986.3 197.61
## - wine_servings      1    1229.08 9241.2 198.57
## - Happiness.Score    1    1386.57 9398.7 199.15
##
## Step:  AIC=194.13
## net_approval ~ Happiness.Score + wine_servings + total_litres_of_pure_alcohol
##
##               Df Sum of Sq    RSS    AIC
## - total_litres_of_pure_alcohol 1      27.24 8136.3 192.24
## <none>                        8109.1 194.13
## + beer_servings      1      96.97 8012.1 195.72
## + spirit_servings     1      17.56 8091.5 196.06
## - wine_servings      1    1188.34 9297.4 196.78
## - Happiness.Score    1    1892.86 10001.9 199.26
##
## Step:  AIC=192.24
## net_approval ~ Happiness.Score + wine_servings
##
```

```
##              Df Sum of Sq      RSS      AIC
## <none>                8136.3 192.24
## + total_litres_of_pure_alcohol  1      27.24  8109.1 194.13
## + beer_servings                1       6.06  8130.3 194.22
## + spirit_servings                1       0.00  8136.3 194.24
## - wine_servings                1    1499.21  9635.5 195.99
## - Happiness.Score              1    1958.76 10095.1 197.58
```

```
step$anova
```

Step	Df	Deviance	Resid. Df	Resid. Dev
	NA	NA	28	7986.322
- spirit_servings	1	25.79172	29	8012.113
- beer_servings	1	96.97033	30	8109.084
- total_litres_of_pure_alcohol	1	27.24260	31	8136.326

Interesting, it seems as though while the other alcohols do not have significant influence in this model, wine consumption

```
library(DAAG)
```

```
## Warning: package 'DAAG' was built under R version 3.3.3
```

```
##
```

```
## Attaching package: 'DAAG'
```

```
## The following object is masked from 'package:MASS':
```

```
##
```

```
## hills
```

```
## The following object is masked from 'package:survival':
```

```
##
```

```
## lung
```

```
TrumpWine <- lm(net_approval ~ Happiness.Score + wine_servings, data=final2)
summary(TrumpWine)
```

```
##
```

```
## Call:
```

```
## lm(formula = net_approval ~ Happiness.Score + wine_servings,
##     data = final2)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -37.218 -11.786   0.555   9.305  38.183
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.10746   18.99607   0.479   0.6350
## Happiness.Score -9.33721    3.41791  -2.732   0.0103 *
## wine_servings  -0.07796    0.03262  -2.390   0.0231 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 16.2 on 31 degrees of freedom
```

```
## Multiple R-squared:  0.4611, Adjusted R-squared:  0.4263
```

```
## F-statistic: 13.26 on 2 and 31 DF,  p-value: 6.9e-05
```

Interesting, the inclusion of wine dampens the estimate of the Happiness Score, but both are still significant and show that the more wine a country drinks and the more happy it is, the less that country approves of Trump. Let's visualize this relationship.

```
library(scatterplot3d)
```

```
## Warning: package 'scatterplot3d' was built under R version 3.3.3
```

```
attach(final2)
```

```
scatterplot3d(Happiness.Score, wine_servings, net_approval, pch=16, highlight.3d=TRUE,  
  type="h" , main="3D Scatterplot")
```

