

# Homework #1: Cantelmo

Robert G. Cantelmo

February 19, 2018

## Section 1: Initial Set Up

```
#Load the Democracy dataset into memory as a dataframe. Use the read.csv function, and the stringsAsFactors = FALSE, and the na.strings = "."
democracy<- read.csv("democracy.csv", stringsAsFactors = FALSE, na.strings = ".")
```

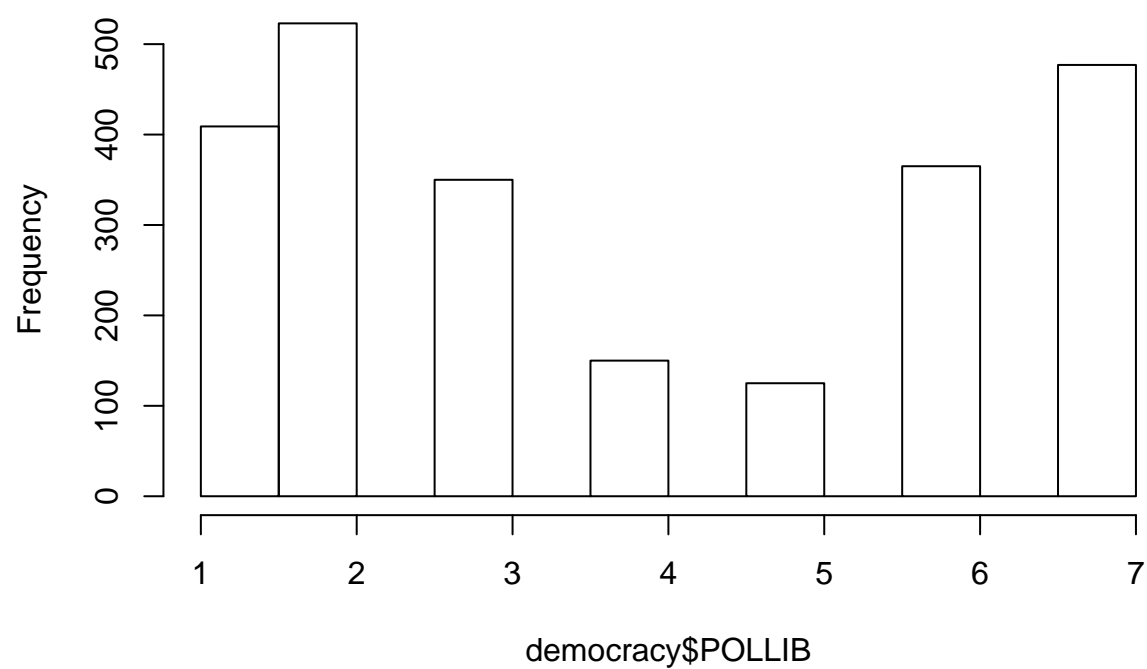
## Section 2: Initial Data Exploration

```
#Report summary statistics (means and medians, at least) for all variables.
summary(democracy)
```

```
##      COUNTRY      CTYNAME      REGION      YEAR
##  Min.   : 1.00    Length:4126    Length:4126    Min.   :1951
## 1st Qu.: 39.00    Class :character    Class :character    1st Qu.:1965
## Median : 71.00    Mode  :character    Mode  :character    Median :1974
## Mean   : 70.18                                Mean   :1973
## 3rd Qu.:103.00                                3rd Qu.:1983
## Max.   :135.00                                Max.   :1990
##
##      BRITCOL      CATH      CIVLIB      EDT
##  Min.   :0.0000    Min.   : 0.00    Min.   :1.000    Min.   : 0.030
## 1st Qu.:0.0000    1st Qu.: 0.00    1st Qu.:2.000    1st Qu.: 2.214
## Median :0.0000    Median :22.25    Median :5.000    Median : 4.540
## Mean   :0.2433    Mean   :37.20    Mean   :4.076    Mean   : 4.853
## 3rd Qu.:0.0000    3rd Qu.:84.00    3rd Qu.:6.000    3rd Qu.: 7.085
## Max.   :1.0000    Max.   :99.00    Max.   :7.000    Max.   :12.810
##                               NA's :1727    NA's :1226
##
##      ELF60      GDPW      MOSLEM      NEWC
##  Min.   :0.0000    Min.   : 480    Min.   : 0.00    Min.   :0.0000
## 1st Qu.:0.1000    1st Qu.:2458    1st Qu.: 0.00    1st Qu.:0.0000
## Median :0.3700    Median :6012    Median : 0.50    Median :0.0000
## Mean   :0.3995    Mean   :8877    Mean   :19.74    Mean   :0.4561
## 3rd Qu.:0.6900    3rd Qu.:12608   3rd Qu.:16.20    3rd Qu.:1.0000
## Max.   :0.9300    Max.   :37903    Max.   :99.90    Max.   :1.0000
## NA's :542
##
##      OIL      POLLIB      REG      STRA
##  Min.   :0.0000    Min.   :1.00    Min.   :0.0000    Min.   :0.0000
## 1st Qu.:0.0000    1st Qu.:2.00    1st Qu.:0.0000    1st Qu.:0.0000
## Median :0.0000    Median :3.00    Median :0.0000    Median :0.0000
## Mean   :0.1001    Mean   :3.86    Mean   :0.3987    Mean   :0.3752
## 3rd Qu.:0.0000    3rd Qu.:6.00    3rd Qu.:1.0000    3rd Qu.:1.0000
## Max.   :1.0000    Max.   :7.00    Max.   :1.0000    Max.   :5.0000
##                               NA's :1727
```

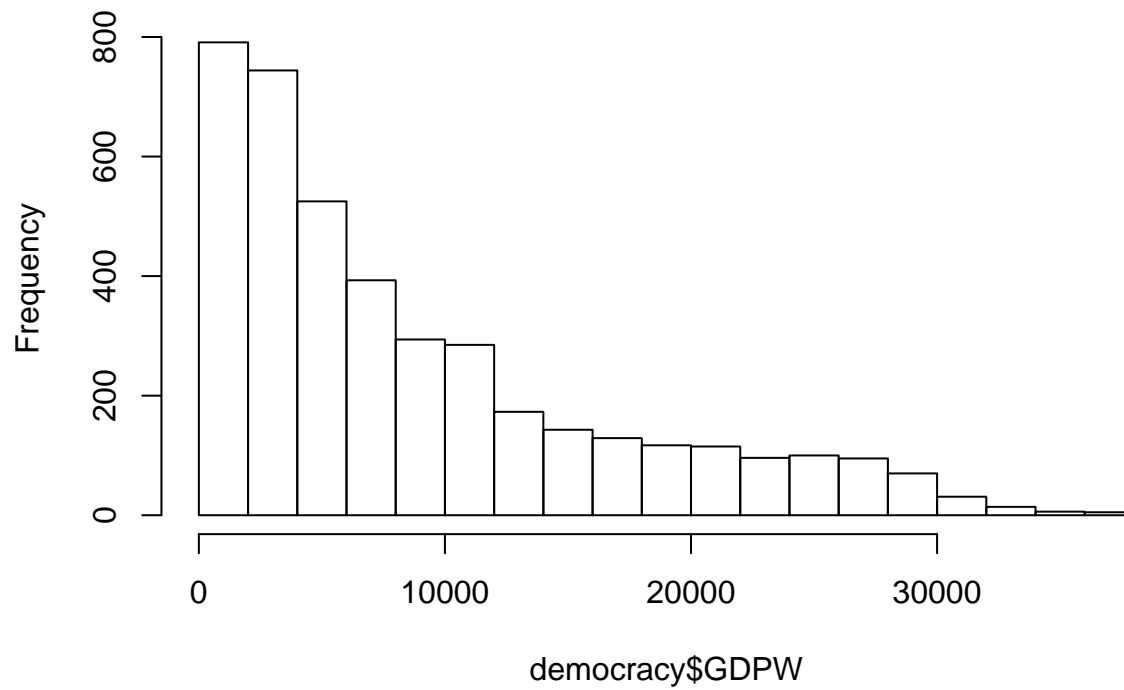
```
#Create a histogram for political liberties in which each unique value of the variable is in its own bin
hist(democracy$POLLIB)
```

**Histogram of democracy\$POLLIB**

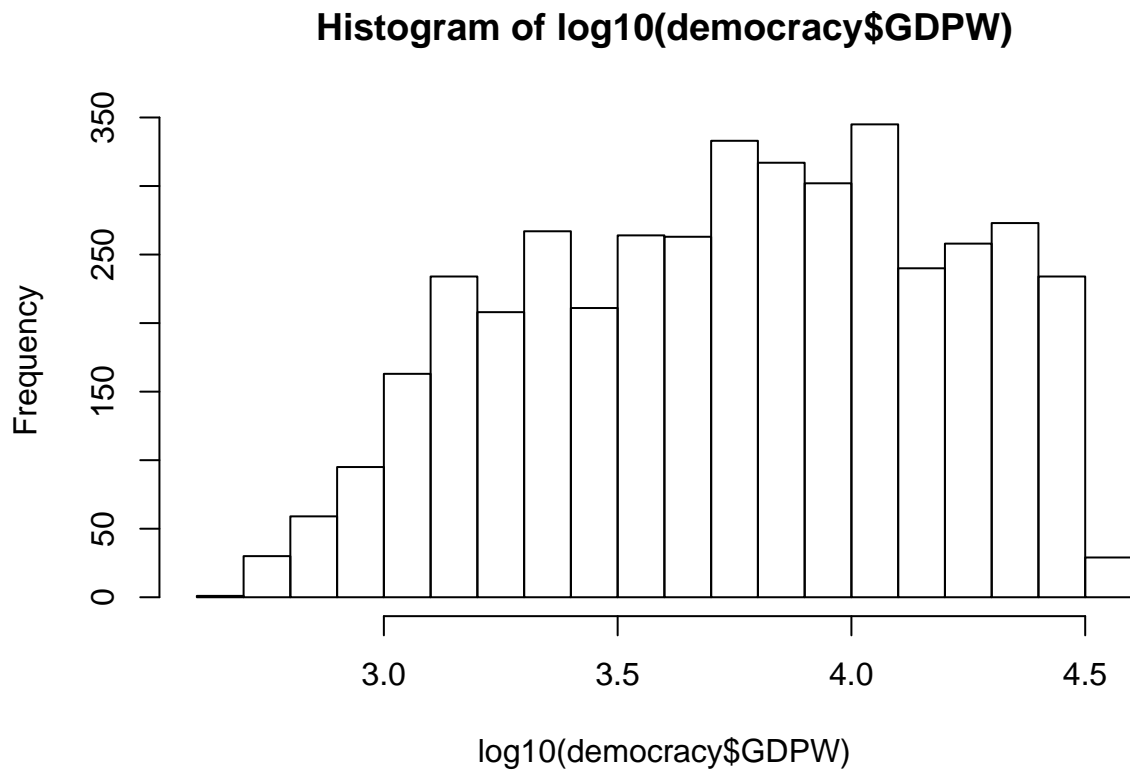


```
#Create a histogram for GDP per capita.  
hist(democracy$GDPW)
```

**Histogram of democracy\$GDPW**



```
#Create a histogram for log GDP per-capita.  
hist(log10(democracy$GDPW))
```



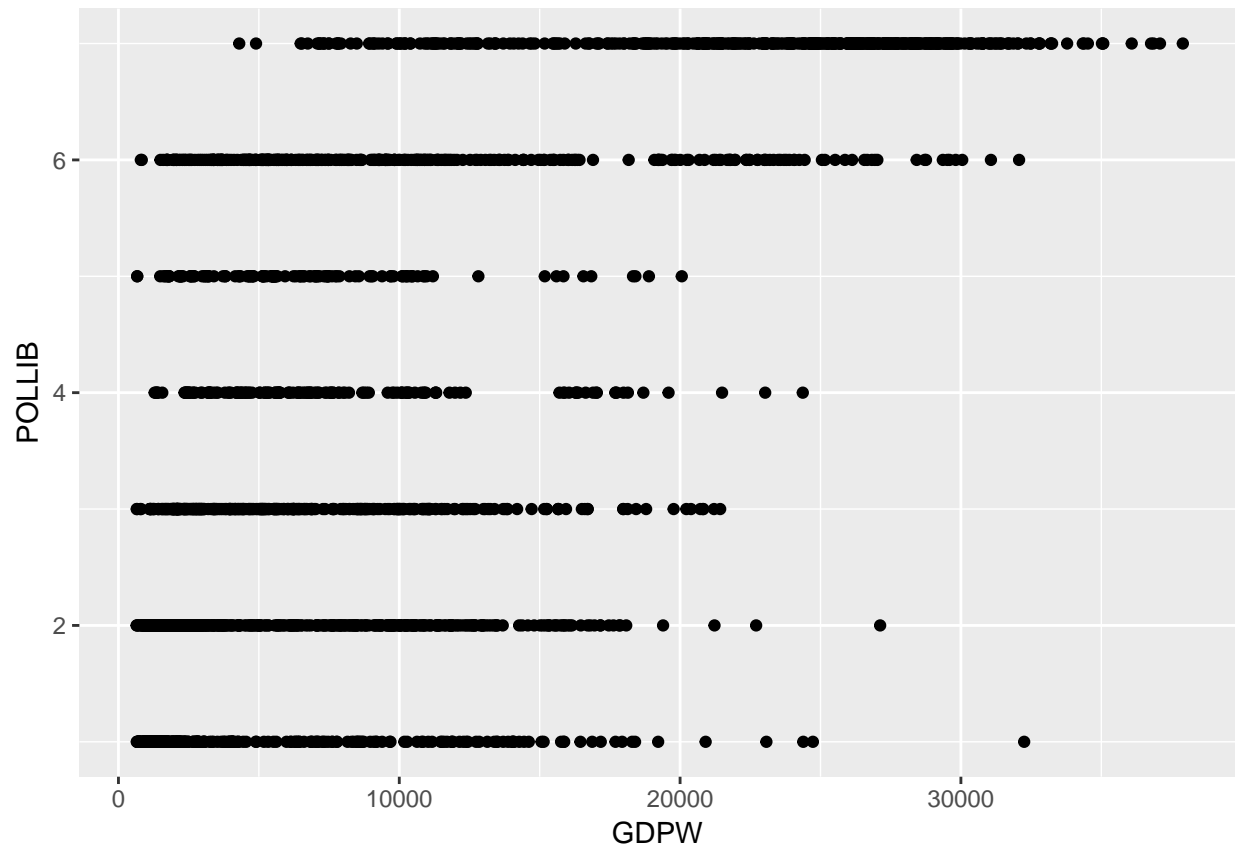
*#How is this histogram different than the one for GDP per capita when it was not logged.*

Section 3: Explore Relationships

*#Create a scatterplot of political liberties against GDP per capita.*

```
ggplot(democracy, aes(x=GDPW, y=POLLIB))+ geom_point()
```

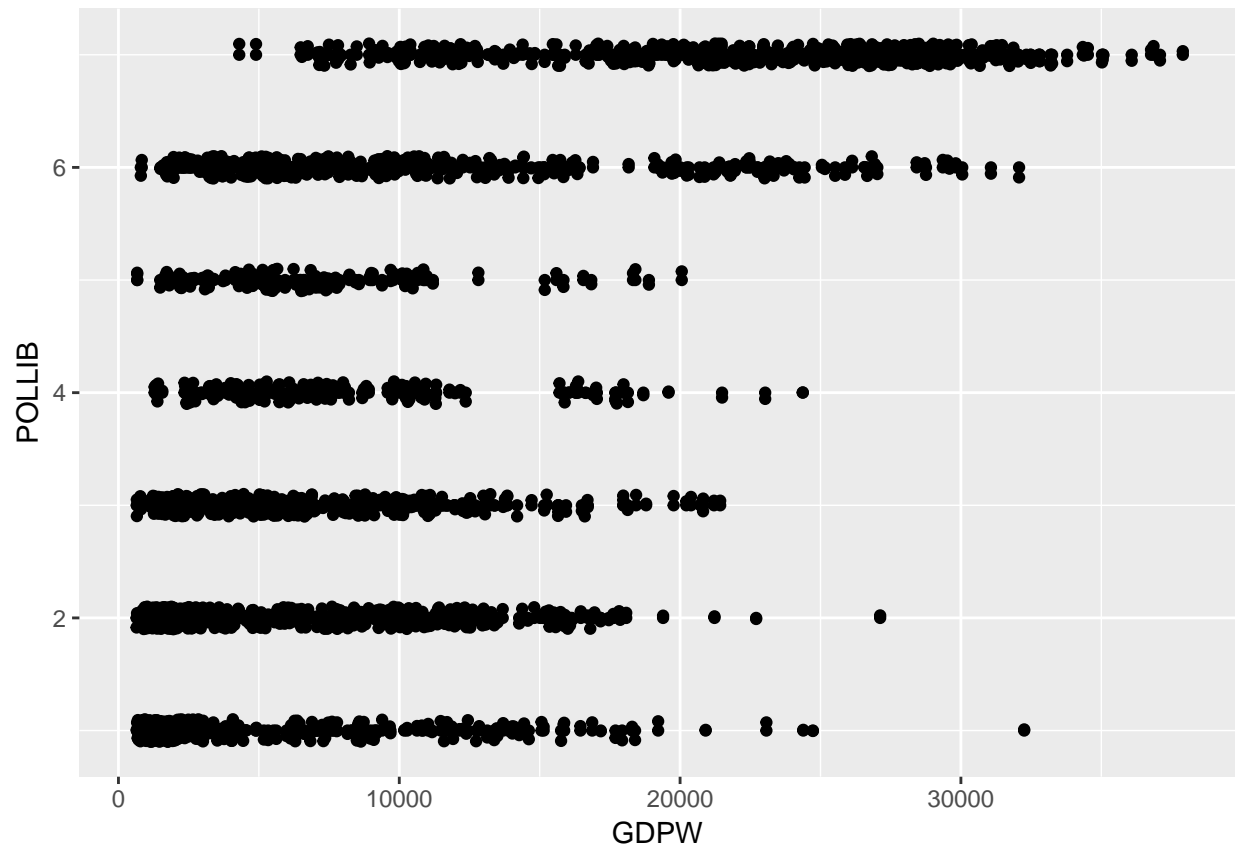
```
## Warning: Removed 1727 rows containing missing values (geom_point).
```



*#When there is a lot of overlap in a scatter plot it is useful to jitter the points (randomly move them)*  
`ggplot(democracy, aes(x=GDPW, y=POLLIB))+ geom_point()+geom_jitter(width = 0, height=.1)`

## Warning: Removed 1727 rows containing missing values (geom\_point).

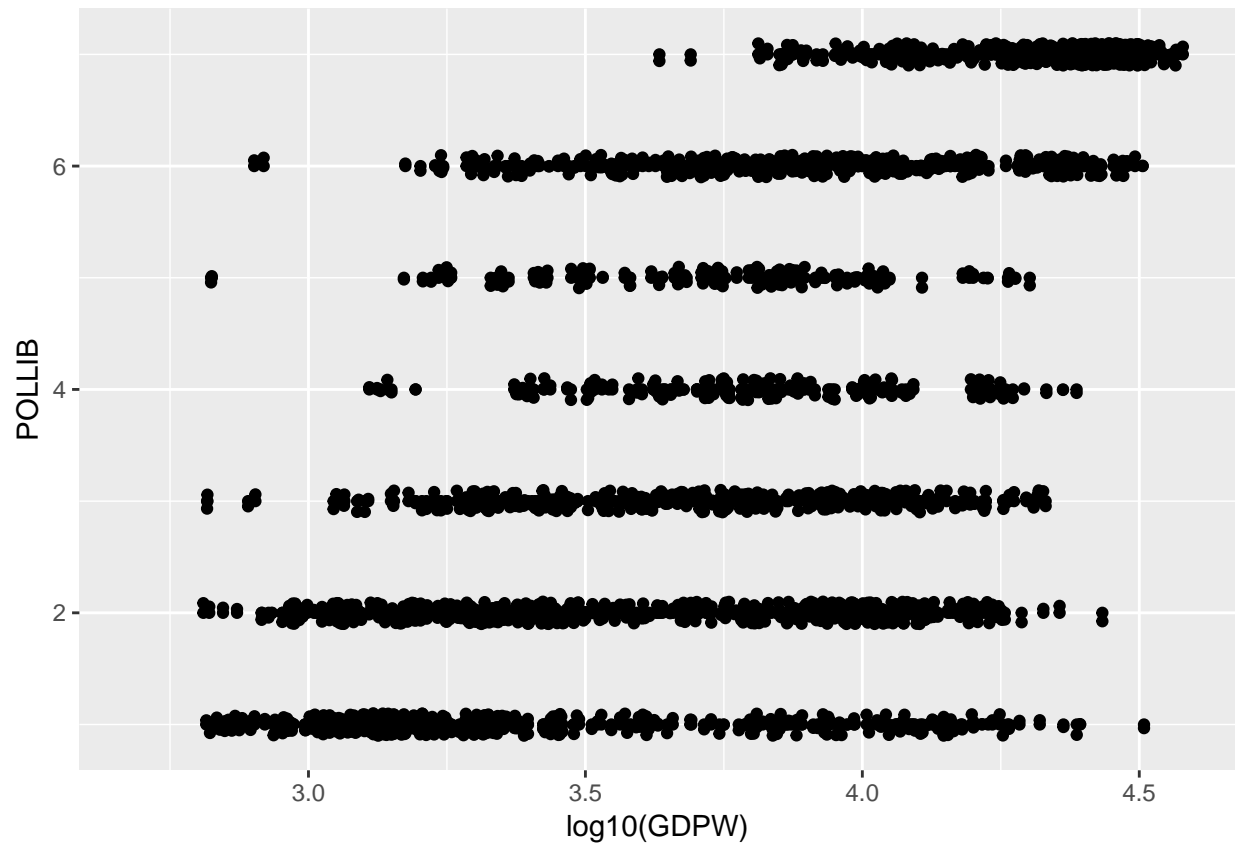
## Warning: Removed 1727 rows containing missing values (geom\_point).



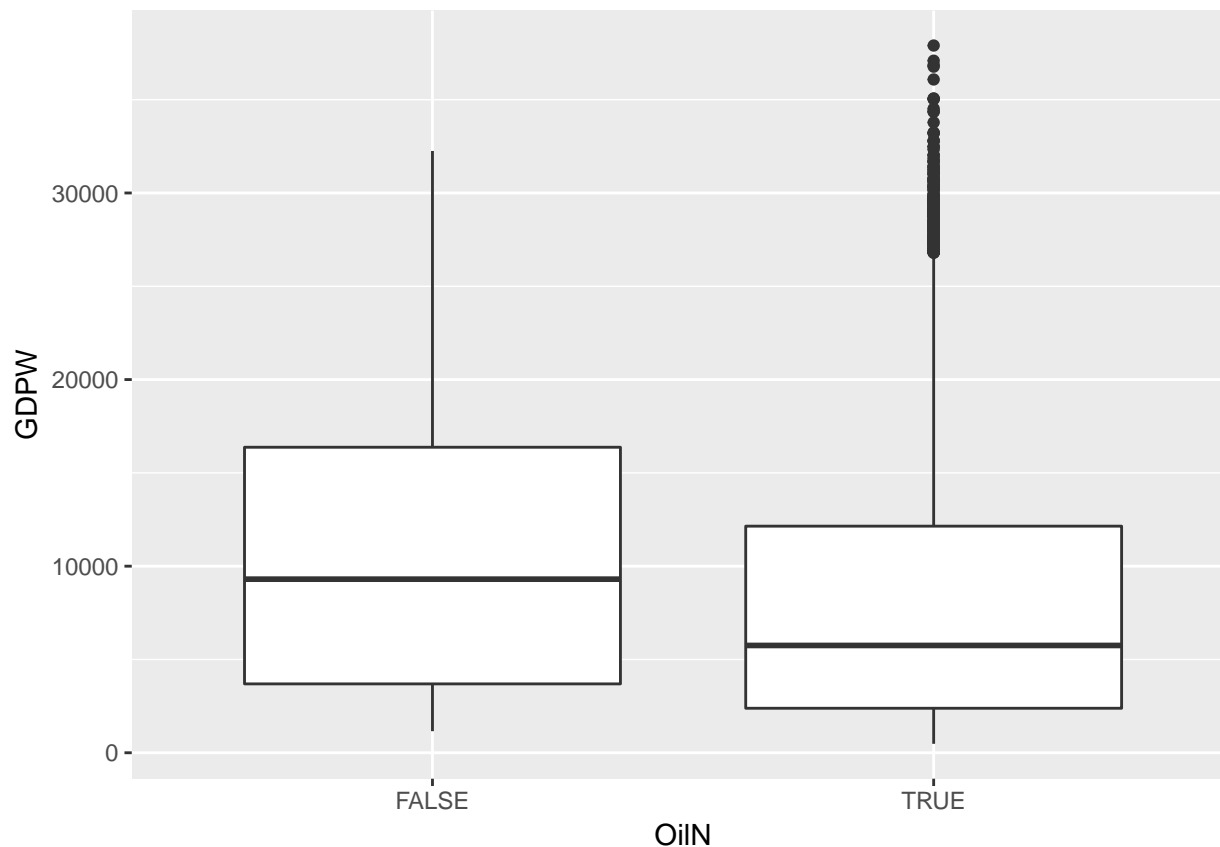
```
#Create a scatterplot of political liberties against log GDP per capita. Jitter the points. How is the
ggplot(democracy, aes(x=log10(GDPW), y=POLLIB))+geom_point()+geom_jitter(width = 0, height=.1)
```

```
## Warning: Removed 1727 rows containing missing values (geom_point).
```

```
## Warning: Removed 1727 rows containing missing values (geom_point).
```



```
#Create a boxplot of GDP per capita for oil producing and non-oil producing nations.
OilN <- c(democracy$OIL==0)
ggplot(democracy, aes(x=OilN , y=GDPW))+ geom_boxplot()
```



#### Section 4: Transform Data and Analyze

*#Calculate the mean GDP per capita in countries with at least 40 percent Catholics. How does it compare*

```
CATHMEAN <- democracy %>% filter(CATH >= 40) %>% select(GDPW) %>% glimpse()
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.3
```

```
## Observations: 1,643
```

```
## Variables: 1
```

```
## $ GDPW <int> 857, 860, 765, 662, 685, 715, 718, 714, 590, 780, 750, 78...
```

```
summary(CATHMEAN)
```

```
##      GDPW
##  Min.   : 590
## 1st Qu.: 4564
##  Median : 8165
##   Mean  :10295
## 3rd Qu.:13889
##   Max.   :37903
```

```
summary(democracy %>% select(GDPW))
```

```
##      GDPW
##  Min.   : 480
## 1st Qu.: 2458
##  Median : 6012
##   Mean  : 8877
## 3rd Qu.:12608
```



```
## Max. :37903
```

```
#The mean GDP per capita for 40% Catholic countries is about $1,400 higher than the world mean.
```

```
#Calculate the average GDP per capita in countries with greater than 60% ethnolinguistic fractionalizat
```

```
VeryHeterog <- democracy %>% filter(ELF60 > .6) %>% select(GDPW)
```

```
LessHeterog <- democracy %>% filter(ELF60 < .6) %>% select(GDPW)
```

```
mean(democracy$GDPW)
```

```
## [1] 8876.959
```

```
UnknownHeterog <- filter(democracy, is.na(ELF60))$GDPW %>% mean(na.rm=TRUE)
```

```
summary(VeryHeterog)
```

```
## GDPW
```

```
## Min. : 505
```

```
## 1st Qu.: 1386
```

```
## Median : 2160
```

```
## Mean : 3591
```

```
## 3rd Qu.: 3751
```

```
## Max. :35069
```

```
#3591
```

```
summary(LessHeterog)
```

```
## GDPW
```

```
## Min. : 480
```

```
## 1st Qu.: 5011
```

```
## Median : 9526
```

```
## Mean :11804
```

```
## 3rd Qu.:17921
```

```
## Max. :37903
```

```
#11804
```

```
summary(UnknownHeterog)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
## 7767 7767 7767 7767 7767 7767
```

```
#7767
```

```
#What was the median of the average years of education in 1985 for all countries?
```

```
AllED1985 <- democracy %>% filter(YEAR==1985) %>% select(EDT)
```

```
summary(AllED1985)
```

```
## EDT
```

```
## Min. : 0.625
```

```
## 1st Qu.: 3.520
```

```
## Median : 5.625
```

```
## Mean : 5.960
```

```
## 3rd Qu.: 7.810
```

```
## Max. :12.810
```

```
## NA's :22
```

```
#5.625
```

```
#Which country was (or countries were) closest to the median years of education in 1985 among all count
```

```
democracy %>% filter(EDT==5.625 & YEAR==1985) %>% glimpse
```

```
## Observations: 1
## Variables: 16
## $ COUNTRY <int> 77
## $ CTYNAME <chr> "Venezuela"
## $ REGION <chr> "South America"
## $ YEAR <int> 1985
## $ BRITCOL <int> 0
## $ CATH <dbl> 95
## $ CIVLIB <int> 2
## $ EDT <dbl> 5.625
## $ ELF60 <dbl> 0.11
## $ GDPW <int> 18362
## $ MOSLEM <dbl> 0
## $ NEWC <int> 0
## $ OIL <int> 1
## $ POLLIB <int> 7
## $ REG <int> 1
## $ STRA <int> 1
```

*#Venezuela*

*#What was the median of the average years of education in 1985 for democracies?*  
ED1985 <- democracy %>% filter(YEAR==1985 & REG==1) %>% select(EDT)  
summary(ED1985)

```
##      EDT
##  Min.   : 1.740
## 1st Qu.: 6.150
##  Median : 7.805
##   Mean   : 8.074
## 3rd Qu.:10.250
##   Max.   :12.810
##  NA's    :6
```

*#7.805*

*#Which democracy was (or democracies were) closest to the median years of education in 1985 among all d*  
democracy %>% filter(YEAR==1985 & EDT==7.805) %>% glimpse

```
## Observations: 1
## Variables: 16
## $ COUNTRY <int> 29
## $ CTYNAME <chr> "Mauritius"
## $ REGION <chr> "Africa"
## $ YEAR <int> 1985
## $ BRITCOL <int> 1
## $ CATH <dbl> 25.7
## $ CIVLIB <int> 2
## $ EDT <dbl> 7.805
## $ ELF60 <dbl> 0.58
## $ GDPW <int> 7474
## $ MOSLEM <dbl> 12.9
## $ NEWC <int> 1
## $ OIL <int> 0
## $ POLLIB <int> 6
## $ REG <int> 1
```

```
## $ STRA      <int> 0
```

```
#Mauritius
```

```
#What were the 25th and 75th percentiles of ethnolinguistic fractionalization for new and old countries
```

```
FracNew <- democracy %>% filter(NEWC==1) %>% select(ELF60)
```

```
FracOld <- democracy %>% filter(NEWC==0) %>% select(ELF60)
```

```
summary(FracNew)
```

```
##      ELF60
```

```
## Min.      :0.040
```

```
## 1st Qu.:0.420
```

```
## Median :0.650
```

```
## Mean      :0.556
```

```
## 3rd Qu.:0.750
```

```
## Max.      :0.930
```

```
## NA's      :251
```

```
# New Countries: 42% @ 25th Percentile & 75% @ 75th Percentile
```

```
summary(FracOld)
```

```
##      ELF60
```

```
## Min.      :0.0000
```

```
## 1st Qu.:0.0600
```

```
## Median :0.1600
```

```
## Mean      :0.2688
```

```
## 3rd Qu.:0.4400
```

```
## Max.      :0.8800
```

```
## NA's      :291
```

```
# Old Countries: 6% @ 25th Percentile & 44% @ 75th Percentile
```