Rcode for POLI502 MidtermPOLI502

Methods of Political Analysis

Student ID: K23040690

2023-10-16

Question 1 Loading Packages and Data

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
## Attaching package: 'gridExtra'
##
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

Question 2 Frequency Table

```
freq_table_oced <- table (world.data$oecd)
ft.oecd <- data.frame(freq_table_oced)
# renames variable
ft.oecd <- ft.oecd %>% rename("OECD Member?" = "Var1")
# adds percentage to the table
ft.oecd $ Percentage <- round(prop.table(ft.oecd $ Freq) * 100, digits = 2)
ft.oecd
```

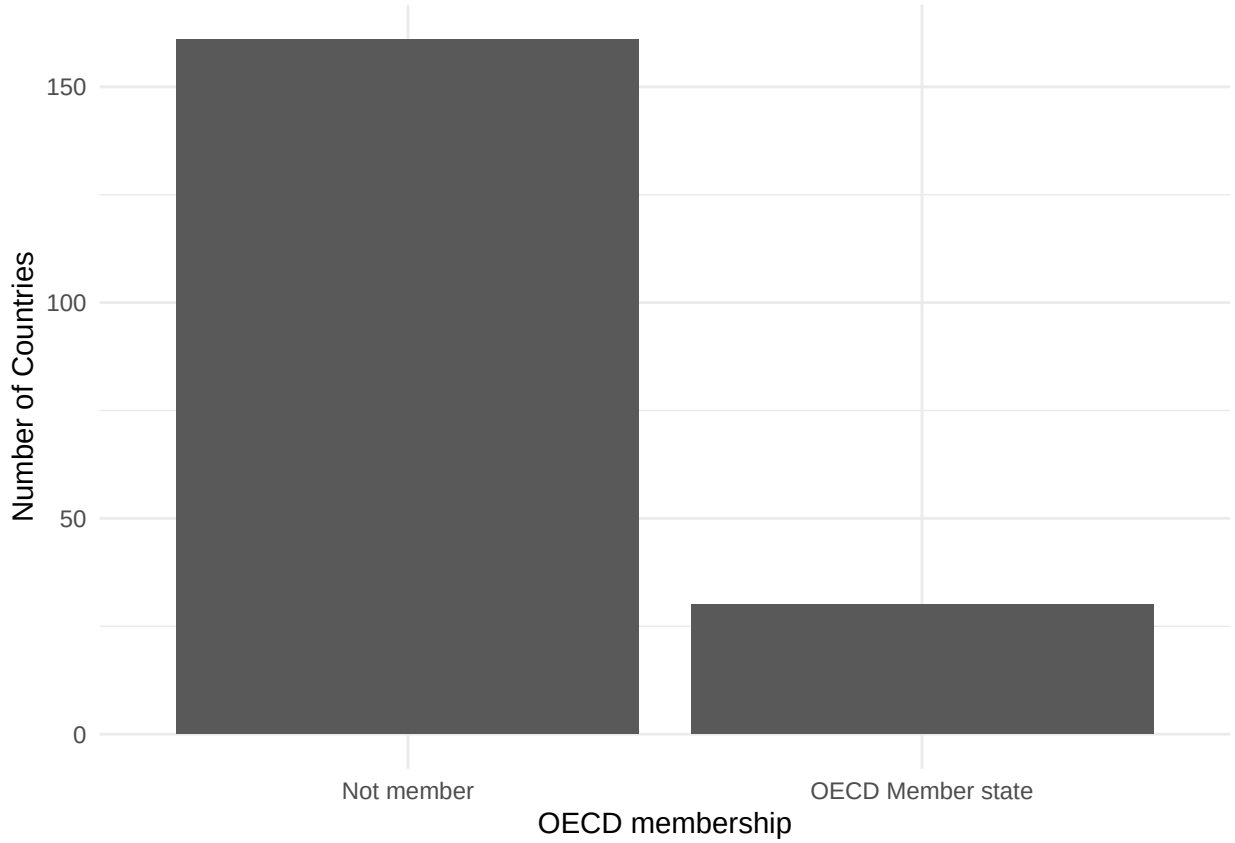
```
##        OECD Member? Freq Percentage
## 1        Not member  161      84.29
## 2 OECD Member state   30      15.71
```

Question 3

3A. 30 countries within the data set are OECD members 3B. 161 countries within the data set are not OECD members 3C. 15.71% of countries within the data set are OECD members, and 3D. 84.29% of countries within the data set are not OECD members.

Question 4 Creating a bar chart of OECD country membership

```
ggplot(world.data, aes (x = oecd)) + geom_bar() +
  labs (x = "OECD membership", y = "Number of Countries") +
  theme_minimal()
```



Question 5 three countries that are OECD members, and three countries that aren't democracies

```
# filters out for OECD members and puts them in a vector
oecd_members <- world.data %>% filter (oecd == "OECD Member state")
# shows the top 3 OECD members
head(oecd_members, 3)
```

```
##     country      colony confidence decentralization dem_other dem_other5
## 1 Australia          UK   46.83889             1.74      58.3 Approx 60%
## 2   Austria       Other   49.68019             1.81     100.0       100%
## 3   Belgium Netherlands   43.31516             1.38     100.0       100%
##   democ_regime    district_size3 durable effectiveness       enpp_3
## 1          Yes     single member      99      90.07092  1-3 parties
## 2          Yes 6 or more members      54      88.88889  4-5 parties
## 3          Yes 6 or more members      55      90.30733 6-11 parties
##                eu fhrate04_rev fhrate08_rev frac_eth frac_eth3 free_business
## 1      Not member    Most free           12   0.0929       Low          90.3
## 2 EU Member state    Most free           12   0.1068       Low          73.6
## 3 EU Member state    Most free           12   0.5554    Medium          92.9
##   free_corrupt free_finance free_fiscal free_govspend free_invest free_labor
## 1           87           90        61.4          64.9          80       94.9
## 2           81           70        51.2          28.8          75       79.1
## 3           73           70        42.2          30.0          80       67.1
##   free_monetary free_overall free_property free_trade gdp08 gdp_10_thou
## 1          82.7         82.6            90       85.1 762.6      2.0822
## 2          79.3         71.6            90       87.5 318.4      2.5356
```

```
## 3          77.9         70.1            80       87.5 369.2      2.3749
##   gdp_cap2 gdp_cap3 gdppcap08 gender_equal3 gini04 gini08   hi_gdp indy
## 1     High     High     35677          High   35.2   35.2 High GDP 1901
## 2     High     High     38152          High   30.0   29.1 High GDP 1156
## 3     High     High     34493          High   25.0   33.0 High GDP 1830
##                oecd  old2006  old2003      pmat12_3    pop03 pop08      pop08_3
## 1 OECD Member state 12.84444 12.49083 High post-mat 19881000  21.0   >=16.8 mil
## 2 OECD Member state 17.03215 16.02893 High post-mat  8090000   8.3 4.4-16.4 mil
## 3 OECD Member state 17.61054 16.74956 High post-mat 10376000  10.7 4.4-16.4 mil
##            popcat3 pr_sys protact3        regime_type3       region sources
## 1 Moderate (1-29m)     No     High Parliamentary democ Asia-Pacific      NA
## 2 Moderate (1-29m)    Yes Moderate Parliamentary democ    W. Europe      NA
## 3 Moderate (1-29m)    Yes     High Parliamentary democ    W. Europe      NA
##          typerel unions urban03 urban06 vi_rel3 votevap00s women05 women09
## 1     Protestant   28.6 91.8942   88.38  20-50%      82.74    24.7    26.7
## 2 Roman Catholic   36.6 67.7750   66.12  20-50%      75.61    33.9    27.9
## 3 Roman Catholic   38.1 97.5114   97.22  20-50%      86.00    34.7    35.3
##   womyear       womyear2  yng2003  young06
## 1    1902 1944 or before 19.97719 19.31490
## 2    1918 1944 or before 16.16753 15.21968
## 3    1919 1944 or before 16.96234 16.64644
```

```r
# filters out the non-democracies
non_democracies <- world.data %>% filter (democ_regime == "No")
# shows the top 3 non-democracies
head(non_democracies, 3)
```

```
##       country   colony confidence decentralization dem_other dem_other5
## 1 Afghanistan       UK         NA               NA      10.5        10%
## 2     Algeria   France   52.05573               NA      40.8 Approx 40%
## 3      Angola Portugal         NA               NA      40.8 Approx 40%
##   democ_regime    district_size3 durable effectiveness enpp_3         eu
## 1           No     single member       4      13.71158        Not member
## 2           No 6 or more members       5      32.62411        Not member
## 3           No                         3      19.14894        Not member
##   fhrate04_rev fhrate08_rev frac_eth frac_eth3 free_business free_corrupt
## 1          2.5            3   0.7693      High            NA           NA
## 2          2.5            3   0.3394    Medium          71.2           32
## 3          2.5            3   0.7867      High          43.4           19
##   free_finance free_fiscal free_govspend free_invest free_labor free_monetary
## 1           NA          NA            NA          NA         NA            NA
## 2           30        83.5          73.4          45       56.4          77.2
## 3           40        85.1          62.8          35       45.2          62.6
##   free_overall free_property free_trade gdp08 gdp_10_thou gdp_cap2 gdp_cap3
## 1           NA            NA         NA  30.6          NA
## 2         56.9            30       70.7 276.0      0.1785      Low   Middle
## 3         48.4            20       70.4 106.3      0.0857      Low   Middle
##   gdppcap08 gender_equal3 gini04 gini08  hi_gdp indy       oecd  old2006
## 1        NA                   NA     NA         1919 Not member       NA
## 2      8033                 35.3   35.3 Low GDP 1962 Not member 4.578136
## 3      5899                   NA     NA Low GDP 1975 Not member 2.450295
##    old2003 pmat12_3    pop03 pop08    pop08_3          popcat3 pr_sys protact3
## 1       NA            NA  27.4 >=16.8 mil Moderate (1-29m)     No
## 2 4.045199          31832610  34.4 >=16.8 mil Moderate (1-29m)    Yes
```

```
## 3 2.930542          13522110  18.0 >=16.8 mil Moderate (1-29m)    Yes
##   regime_type3      region sources        typerel unions urban03 urban06
## 1 Dictatorship Middle East      NA         Muslim     NA      NA   23.28
## 2 Dictatorship      Africa      NA         Muslim     NA 58.8302   63.94
## 3 Dictatorship      Africa      NA Roman Catholic     NA 36.1806   53.96
##   vi_rel3 votevap00s women05 women09 womyear   womyear2  yng2003  young06
## 1                NA      NA    27.7      NA                 NA       NA
## 2    >50%        NA      NA     7.7    1962 After 1944 33.91887 28.94154
## 3                NA      NA    37.3    1975 After 1944 47.62524 46.32196
```

Question 6 statistics on GDP

```
# summary provides the range, median, mean, and 1st and 3rd quartile
summary(world.data$gdp_10_thou)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.0090  0.0503  0.1897  0.6018  0.6320  4.7354      14
```

```
# creating a vector of gdp with no NAs to determine the standard deviation
gdp_no_na <- na.omit(world.data$gdp_10_thou)
sd(gdp_no_na)
```

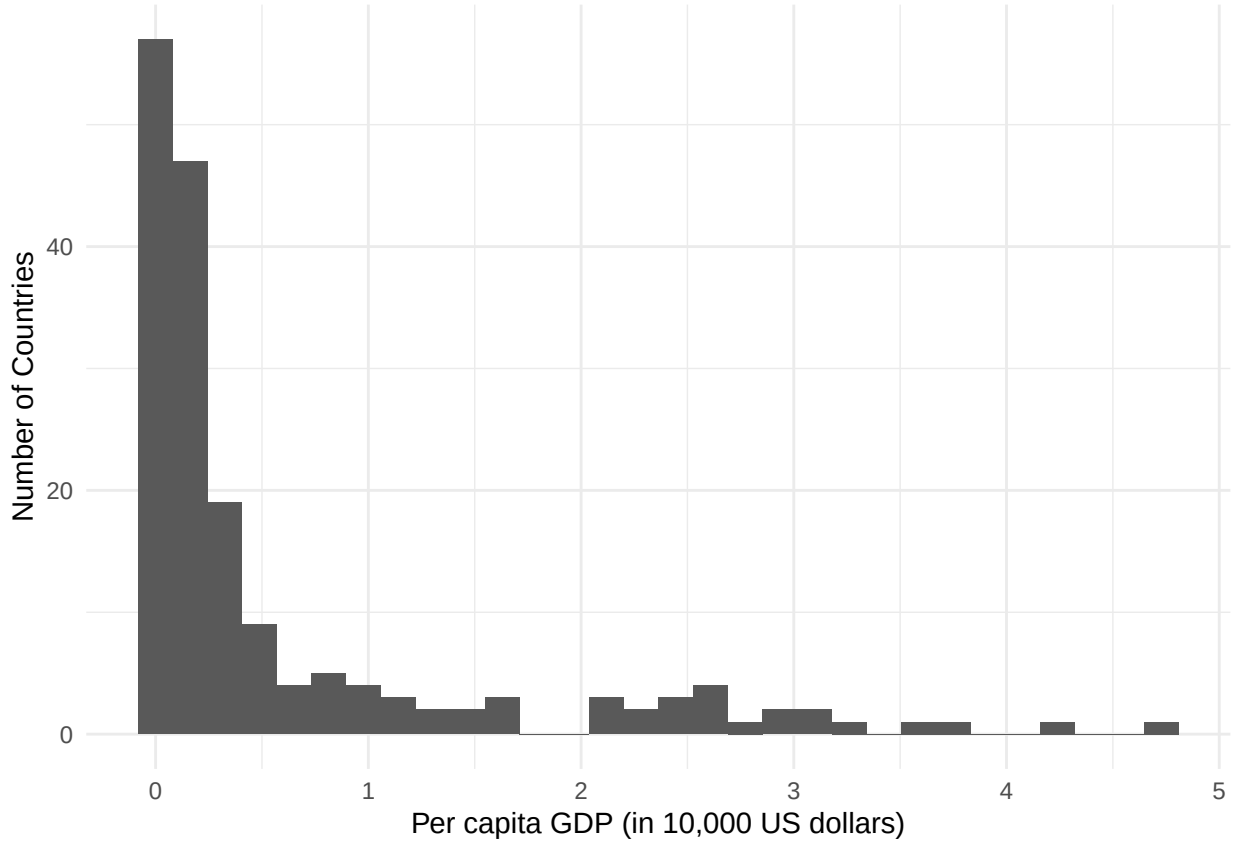
```
## [1] 0.9433982
```

Question 7 Since the median is lower than the mean, the distribution will have a positive skew. That is, the majority of case will occur on the left side of the graph, but there will be a longer tail on the right side of the graph that pulls the mean up.

Question 8 Graphing GDP

```
gdp_data_frame <- data.frame (gdp_no_na)
ggplot (gdp_data_frame, aes (x = gdp_no_na)) + geom_histogram() +
  labs (x = "Per capita GDP (in 10,000 US dollars)", y = "Number of Countries") +
  theme_minimal()
```

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



Question 9 Listing 2 countries with gdp per capita over 40k

```
# filter out countries to just have the countries that are at 40k or higher
gdp_40k <- world.data %>% filter (gdp_10_thou >= 4)
head(gdp_40k)
```

```
##      country      colony confidence decentralization dem_other dem_other5
## 1 Luxembourg Netherlands         NA             1.16       100       100%
## 2     Norway        none   66.45381             1.79       100       100%
##   democ_regime    district_size3 durable effectiveness       enpp_3
## 1          Yes 6 or more members      NA      96.92672  4-5 parties
## 2          Yes 6 or more members      54      90.07092 6-11 parties
##                eu fhrate04_rev fhrate08_rev frac_eth frac_eth3 free_business
## 1 EU Member state    Most free           12   0.5302    Medium          75.1
## 2      Not member    Most free           12   0.0586       Low          88.8
##   free_corrupt free_finance free_fiscal free_govspend free_invest free_labor
## 1           83           80        65.9          58.5          95       40.4
## 2           79           60        50.5          49.8          65       47.1
##   free_monetary free_overall free_property free_trade gdp08 gdp_10_thou
## 1          78.9         75.4            90       87.5  38.3      4.7354
## 2          74.2         69.4            90       89.2 277.3      4.1974
##   gdp_cap2 gdp_cap3 gdppcap08 gender_equal3 gini04 gini08   hi_gdp indy
## 1     High     High     78599                30.8   31.0 High GDP 1839
## 2     High     High     58138          High  25.8   25.8 High GDP 1905
##                oecd  old2006  old2003          pmat12_3   pop03 pop08
## 1 OECD Member state 13.79791 14.92056                    448000   0.5
```

```
## 2 OECD Member state 15.03641 14.87521 Moderate post-mat 4562000   4.8
##        pop08_3          popcat3 pr_sys protact3        regime_type3      region
## 1    <=4.3 mil Small (under 1m)    Yes          Parliamentary democ   W. Europe
## 2 4.4-16.4 mil Moderate (1-29m)    Yes     High Parliamentary democ Scandinavia
##   sources        typerel unions urban03 urban06 vi_rel3 votevap00s women05
## 1      NA Roman Catholic   39.5 92.4990   82.68    <20%      56.50    23.3
## 2      NA     Protestant   51.7 75.5492   77.50    <20%      76.54    38.2
##   women09 womyear       womyear2  yng2003  young06
## 1    20.0    1919 1944 or before 18.94910 18.81388
## 2    38.2    1907 1944 or before 19.73542 19.38013
```

Question 10 Calculating Standard Error Standard Error = standard deviaion / sqrt of n

```
n <- length(gdp_no_na)
s <- sd(gdp_no_na)
s/sqrt(n)
```

```
## [1] 0.07091015
```

The standard error is 0.0709

Question 11 Calculating the 95% confidence interval for GDP capita

```
# placing the mean in a variable for easy use
mean_gdp_capita <- mean(gdp_no_na)
mean_gdp_capita
```

```
## [1] 0.6018186
```

```
# placing the standard error in a variable for easy use
standard_error <- s/sqrt(n)
standard_error *2
```

```
## [1] 0.1418203
```

```
# Upper bound of the confidence interval
mean_gdp_capita + (standard_error *2)
```

```
## [1] 0.7436389
```

```
# Lower bound of the confidence interval
mean_gdp_capita - (standard_error *2)
```

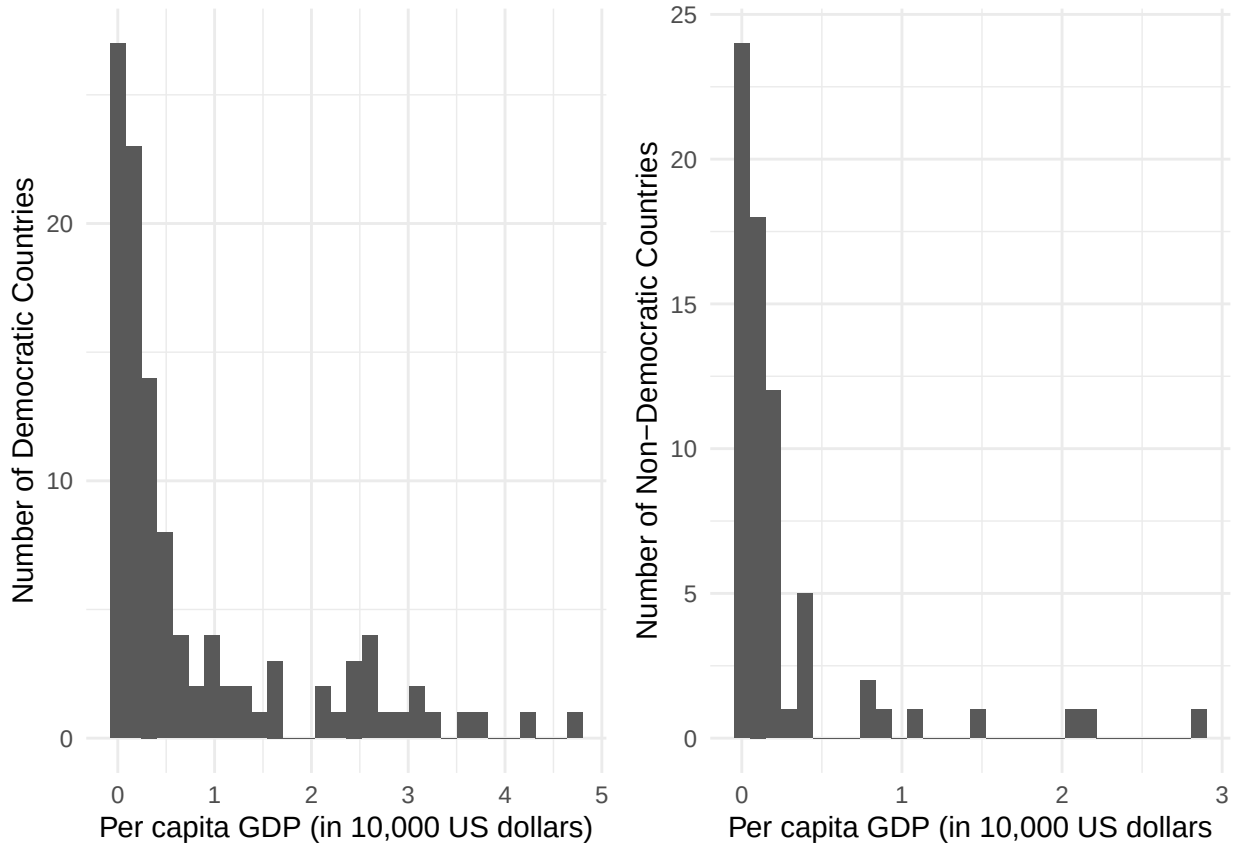
```
## [1] 0.4599983
```

the confidence interval is 0.6 plus or minus 0.14 yielding a range from 0.46 to 0.74. It should be remembered that this variable is in units of 10,000, meaning that the confidence interval is $4,600 - $7,400

Question 12 Graphing Gdp per capita by whether democracy or not

```r
# separating democracies and non-democracies
gdp_cap_demo <- world.data %>% filter (democ_regime == "Yes") %>% select (gdp_10_thou)
dp_cap_no_demo <-world.data %>% filter (democ_regime == "No") %>% select (gdp_10_thou)

# putting these into a data frame to use ggplot
demo_gdp_frame <- data.frame (gdp_cap_demo)
no_demo_gdp_frame <- data.frame (dp_cap_no_demo)

# creating ggplots of each
demo_plot <- ggplot(demo_gdp_frame, aes(x = gdp_10_thou)) + geom_histogram() +
  labs (x = "Per capita GDP (in 10,000 US dollars)", y = "Number of Democratic Countries") +
  theme_minimal()

non_demo_plot <- ggplot(no_demo_gdp_frame, aes(x = gdp_10_thou)) + geom_histogram() +
  labs (x = "Per capita GDP (in 10,000 US dollars", y = "Number of Non-Democratic Countries") + theme_m

# putting the plots side by side
grid.arrange(demo_plot,non_demo_plot,ncol=2)
```

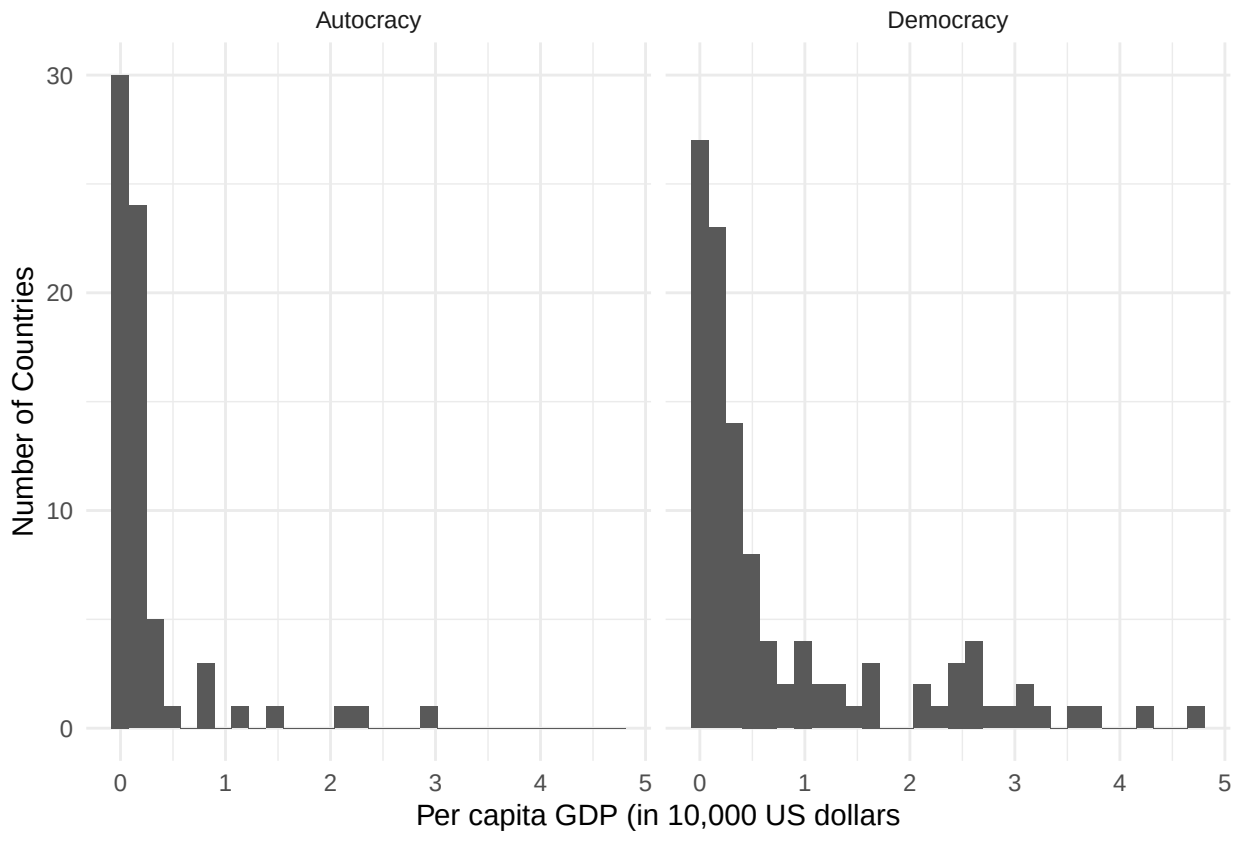
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 5 rows containing non-finite values (`stat_bin()`).

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

## Warning: Removed 7 rows containing non-finite values (`stat_bin()`).
```



Question 13 Regraphing the graphs above

```
# creating new isolate data frame
dem_data <- world.data[c("democ_regime", "gdp_10_thou")]

# removing NAs from democratic data
dem.gdp <- na.omit(dem_data)
data_frame_dem_gdp <- data.frame(dem.gdp)

# Changing yes and no to be democracy and autocracy
dem.gdp$democ_regime <- ifelse (dem.gdp$democ_regime == "Yes", "Democracy", "Autocracy")

# Creating the dual plots
g <- ggplot (dem.gdp, aes (x = gdp_10_thou))
g <- g + geom_histogram()
g <- g + xlab("Per capita GDP (in 10,000 US dollars")
g <- g + ylab ("Number of Countries")
g <- g + theme_minimal()
g <- g + facet_wrap (~ democ_regime)
g
```

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



Question 14 Calculating the confidence interval for democracies

```r
# Calculating the mean and standard deviation
mean_gdp_capita_demo <- mean(subset (dem.gdp, democ_regime == "Democracy")$gdp_10_thou)
sd_gdp_capita_demo <- sd(subset(dem.gdp, democ_regime == "Democracy")$gdp_10_thou)

# Finding the n to determine the standard error
n_demo <- length (subset(dem.gdp, democ_regime == "Democracy")$gdp_10_thou)

# Calculating the standard error
standard_error_demo <- sd_gdp_capita_demo / sqrt(n_demo)

# 95% confidence interval
#Upper Bound
mean_gdp_capita_demo + (standard_error_demo *2)
```

```
## [1] 1.008499
```

```r
# Lower Bound
mean_gdp_capita_demo - (standard_error_demo *2)
```

```
## [1] 0.5942867
```

Democracies have a mean per capita GDP of 8,014 plus or minus $2,071, leaving a 95% confidence interval of $5,943 to $10,084 (numbers are rounded)

Question 15 Calculating the confidence interval for autocracies

```r
# Calculating the mean and standard deviation
mean_gdp_capita_auto <- mean(subset (dem.gdp, democ_regime == "Autocracy")$gdp_10_thou)
sd_gdp_capita_auto <- sd(subset(dem.gdp, democ_regime == "Autocracy")$gdp_10_thou)

# Finding the n to determine the standard error
n_auto <- length (subset(dem.gdp, democ_regime == "Autocracy")$gdp_10_thou)

# Calculating the standard error
standard_error_auto <- sd_gdp_capita_auto / sqrt(n_auto)
standard_error_auto *2
```

```
## [1] 0.129515
```

```r
# 95% confidence interval
#Upper Bound
mean_gdp_capita_auto + (standard_error_auto *2)
```

```
## [1] 0.4114282
```

```r
# Lower Bound
mean_gdp_capita_auto - (standard_error_auto *2)
```

```
## [1] 0.1523983
```

Autocracies have a mean gdp/capita of $2,819, plus or mins $1,295 to construct a 95% confidence interval. This leaves a range of $4,114 to $1,523

Since there is no cross over between the two confidence intervals, we can be very confidence that democracies do have higher GDPs/capita on average than autocracies.

Question 16 Determining the probability of rain after clouds.

```
# first determine the probability of just clouds.
# c = clouds
# r = rain
# Using the law of total probability to find the probability of clouds are any
  # day
# p(c) = p(c|r) *p(r) + p(c|~r)*p(~r)
# p(c|r) = 95%
# p(r) = 30%
# p(c|~r) = 25%
# p(r) = 100 - 30 = 70%

prob_clouds <- (.95*.3) + (.25*.7)
prob_clouds
```

```
## [1] 0.46
```

```
# prob cloud = 46%

# Using Bayes' rule to determine p(r|c)
# p(r|c) = (p(c|r)*p(r)) / p(c)

prob_rain_given_clouds <- (.95*.3)/.46
prob_rain_given_clouds
```

```
## [1] 0.6195652
```

The probability of seeing rain after seeing clouds is fairly high at 61.96%

Question 17 Multiple Bayesian Questions

17a.

```
# using the formulas given within the question, the mean is detemined by
  #a/(a+b) and the standard deviation by ab/ (a+b+1)(a+b)^2
# a and b are both equal to 1.5

prior_mean <- 1.5/(1.5+1.5)
prior_standard_d <- (1.5*1.5)/(((1.5+1.5+1)*(1.5+1.5))^2)
prior_mean
```

```
## [1] 0.5
```

```
prior_standard_d
```

```
## [1] 0.015625
```

The prior mean is 0.5 and the prior standard devitation is 0.0156

17b.

```
# Again, using the formula given:
pbeta(.6, 1.5, 1.5)
```

```
## [1] 0.62647
```

The probability of the prior theta being greater than 60% is 62.6%

17c.

```
# Using the idea that theta is equal to 60%, we use the formula given.
likelihood <- (.6^37)*((1-.6)^13)
likelihood
```

```
## [1] 4.153136e-14
```

The likelihood of the data given the prior appears to be very small (though I'm not sure what is a typical range for likelihoods) at 4.15e-14

17d.

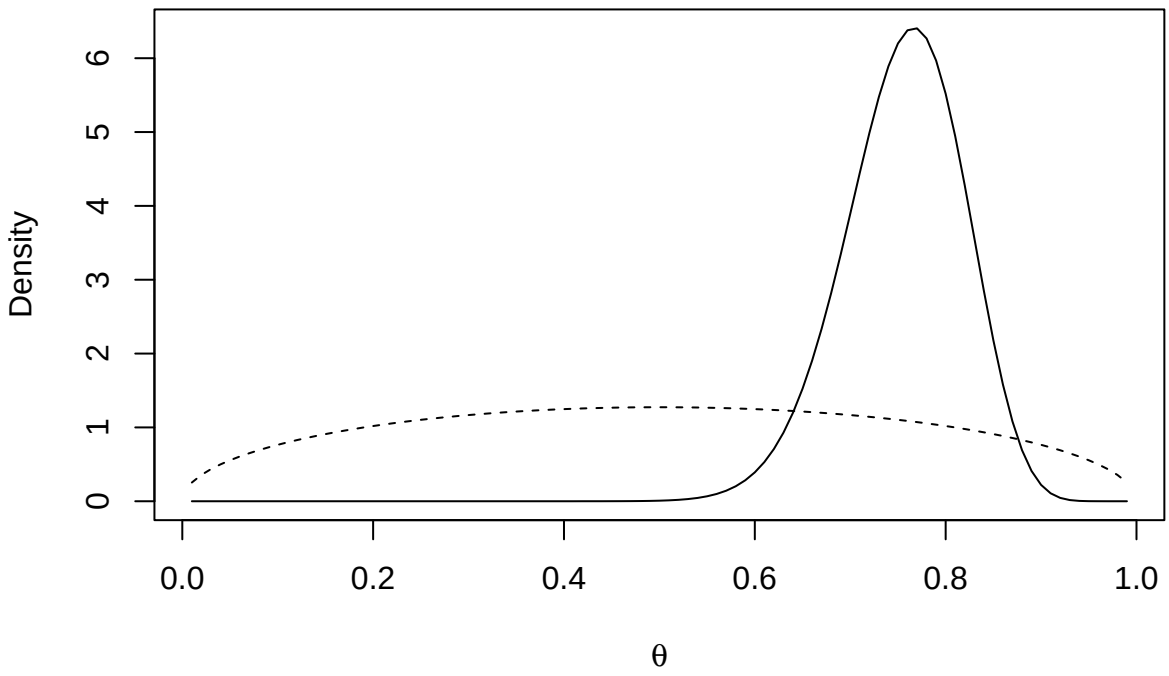
```
# the prior distribution
prior_dist <- (.6^.5)*(.4^.5)
posterior_dist <- likelihood * prior_dist
posterior_dist
```

```
## [1] 2.034613e-14
```

Again, without knowing a typical value to judge by, the posterior distribution is 2.034613e-14

17e.

```
theta<-seq(0.01,0.99,0.01)
a <- 1.5
b <- 1.5
c <- 37 - 1.5
d <- 13 - 1.5
 prior<-dbeta(theta,a,b)
 posterior<-dbeta(theta,c,d)
 plot(theta,posterior,xlab=expression(theta),ylab="Density",type="l")+
 lines(theta, prior,lty=2)
```




```
## integer(0)
```

17f.

The graph above is showing both the prior and posterior distributions. The prior is the dotted line, the posterior is the solid line. The posterior shows that the distribution we have thus far has more approval of the professor than initially anticipated