

Tutorial 9: Data Management

Jan Vogler (jan.vogler@duke.edu)

October 23, 2015

Today's Agenda

1. Data management I: reading data, keeping/deleting variables
2. Data management II: data transformation
3. Data management III: creating new variables
4. Some more useful R commands for data management
5. The most important R packages for applied work

1. Data management I: reading data, keeping/deleting variables

Don't forget that in most cases you will need to

R can save its own datafiles. Those will have the format ".Rdata". In order to load such a dataset, you can simply use the load command.

```
load(filename.Rdata)
```

```
## Error in load(filename.Rdata): object 'filename.Rdata' not found
```

R can also natively read some other file formats, including .csv and .txt files.

```
### .csv files
```

```
csvdata = read.csv("filename.csv", stringsAsFactors = FALSE)
```

```
## Warning in file(file, "rt"): cannot open file 'filename.csv': No such file  
## or directory
```

```
## Error in file(file, "rt"): cannot open the connection
```

```
# stringsAsFactors=FALSE is important because otherwise R will load  
# character variables as factors, treating them as numerical under the  
# surfae, which can lead to complications later on
```

```
### .txt files
```

```
textdata = read.table("filename.txt")
```

```
## Warning in file(file, "rt"): cannot open file 'filename.txt': No such file  
## or directory
```

```
## Error in file(file, "rt"): cannot open the connection
```

Additionally, there are many other data formats. Some of them require the foreign package.

```
library(foreign)
spssdata = read.spss("filename", to.data.frame = TRUE)
```

```
## Error in read.spss("filename", to.data.frame = TRUE): unable to open file: 'No such file or directory'
```

```
setwd("C:/Users/Jan/OneDrive/Documents/GitHub/ps630_lab/")
library(foreign)
LDC = read.dta("LDC_IO_replication.dta")
```

For data files that were saved by the most recent version of STATA (STATA 13), you will need another package called “readstata13”. Please use the following command to install it: `install.packages(“readstata13”)`

```
read.dta13("filename.dta")
```

```
## Error in eval(expr, envir, enclos): could not find function "read.dta13"
```

It is most likely that you will need to use more than one dataset for your empirical analysis. You can merge two datasets by using the merge command.

```
merge(dataset1, dataset2, by = c("Country", "Year"))
```

```
## Error in merge(dataset1, dataset2, by = c("Country", "Year")): object 'dataset1' not found
```

To delete data, we simply use the following command:

```
LDC$ecris2=NULL
```

We can also only keep some data that we really need:

```
LDC=LDC[,c("ctylabel", "date", "polityiv_update2", "gdp_pc_95d", "newtar", "l1polity", "l1polity", "l1signed",
```

The following command allows us to only look at cases on which we have all observations of specific variables.

```
complete = with(LDC, complete.cases(polityiv_update2,gdp_pc_95d))
# The "with" command allows you to evaluate a file for certain expressions, such as complete.cases
LDC=LDC[complete,]
# Then we subset the file and only take the complete cases
```

Before doing any analysis, you should inspect your variables closely. Make sure that you are aware of the properties of your most important variables.

```
summary(LDC$polityiv_update2)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -10.000  -7.000   -4.000  -1.121   7.000   10.000
```

```
summary(LDC$gdp_pc_95d)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    91.02  380.70   978.00 2320.00 2437.00 44160.00
```

```
summary(LDC$newtar)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     0.10  11.70   17.80   21.73   27.50  102.20    2014
```

2. Data management II: data transformation

We can transform variables in a number of ways. One of the most common ways to transform data is to square it to estimate curvilinear relationships. Let us construct a squared version of the Polity IV Score.

```
LDC$polityiv_squared=(LDC$polityiv_update2)^2
LDC$l1polity_squared=(LDC$l1polity)^2
```

Let us see whether there is a curvilinear relationship between the Polity IV Score and tariff levels.

```
main=lm(newtar ~ l1polity + l1polity_squared + l1signed + l1office + l1gdp_pc + l1lnpop + l1ecris2 + l1l1bpc1 + l1avnewtar +
summary(main)
```

```
##
## Call:
## lm(formula = newtar ~ l1polity + l1polity_squared + l1signed +
##      l1office + l1gdp_pc + l1lnpop + l1ecris2 + l1bpc1 + l1avnewtar +
##      factor(ctylabel) - 1, data = LDC)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -34.020  -3.431  -0.043   3.154  34.532
##
## Coefficients:
##              Estimate Std. Error t value
## l1polity          -5.050e-01  1.025e-01  -4.925
## l1polity_squared  -5.775e-02  2.050e-02  -2.817
## l1signed           5.252e-01  7.520e-01   0.698
## l1office          -1.922e-01  6.913e-02  -2.781
## l1gdp_pc          -5.078e-05  4.009e-04  -0.127
## l1lnpop           -2.090e+01  3.660e+00  -5.712
## l1ecris2          -1.582e+00  1.131e+00  -1.399
## l1bpc1             2.350e+00  8.875e-01   2.648
## l1avnewtar         3.481e-01  6.868e-02   5.068
## factor(ctylabel)Albania      3.250e+02  5.582e+01   5.823
## factor(ctylabel)Algeria      3.738e+02  6.344e+01   5.892
## factor(ctylabel)Argentina    3.799e+02  6.408e+01   5.928
## factor(ctylabel)Bahrain      2.719e+02  4.958e+01   5.483
## factor(ctylabel)Bangladesh   4.454e+02  6.862e+01   6.491
```

## factor(ctylabel)Belarus	3.439e+02	6.025e+01	5.709
## factor(ctylabel)Benin	3.487e+02	5.715e+01	6.102
## factor(ctylabel)Bolivia	3.432e+02	5.840e+01	5.877
## factor(ctylabel)Botswana	3.231e+02	5.277e+01	6.123
## factor(ctylabel)Brazil	4.223e+02	6.943e+01	6.082
## factor(ctylabel)Bulgaria	3.518e+02	5.909e+01	5.955
## factor(ctylabel)Burundi	3.486e+02	5.789e+01	6.022
## factor(ctylabel)Cambodia	3.663e+02	6.007e+01	6.098
## factor(ctylabel)Cameroon	3.555e+02	6.028e+01	5.898
## factor(ctylabel)CentralAfricanRepublic	3.284e+02	5.553e+01	5.914
## factor(ctylabel)Chad	3.391e+02	5.844e+01	5.803
## factor(ctylabel)Chile	3.556e+02	6.085e+01	5.844
## factor(ctylabel)China	4.620e+02	7.762e+01	5.951
## factor(ctylabel)Colombia	3.863e+02	6.439e+01	6.000
## factor(ctylabel)Congo	3.232e+02	5.456e+01	5.924
## factor(ctylabel)CostaRica	3.290e+02	5.560e+01	5.918
## factor(ctylabel)Coted'Ivoire	3.571e+02	6.047e+01	5.907
## factor(ctylabel)Cyprus	2.978e+02	5.065e+01	5.879
## factor(ctylabel)DominicanRepublic	3.468e+02	5.862e+01	5.917
## factor(ctylabel)Ecuador	3.593e+02	6.000e+01	5.989
## factor(ctylabel)Egypt	3.999e+02	6.587e+01	6.071
## factor(ctylabel)ElSalvador	3.357e+02	5.736e+01	5.853
## factor(ctylabel)Estonia	2.986e+02	5.268e+01	5.668
## factor(ctylabel)Ethiopia	3.890e+02	6.641e+01	5.858
## factor(ctylabel)Fiji	2.948e+02	5.059e+01	5.827
## factor(ctylabel)Gabon	3.075e+02	5.155e+01	5.965
## factor(ctylabel)Gambia	2.989e+02	5.232e+01	5.714
## factor(ctylabel)Ghana	3.559e+02	6.132e+01	5.804
## factor(ctylabel)Guatemala	3.453e+02	5.934e+01	5.818
## factor(ctylabel)Guinea	3.348e+02	5.809e+01	5.763
## factor(ctylabel)Guyana	2.966e+02	5.073e+01	5.846
## factor(ctylabel)Haiti	3.414e+02	5.845e+01	5.841
## factor(ctylabel)Honduras	3.306e+02	5.778e+01	5.721
## factor(ctylabel)Hungary	3.532e+02	6.015e+01	5.872
## factor(ctylabel)India	4.914e+02	7.606e+01	6.460
## factor(ctylabel)Indonesia	4.135e+02	7.062e+01	5.855
## factor(ctylabel)Israel	3.289e+02	5.717e+01	5.753
## factor(ctylabel)Jamaica	3.247e+02	5.465e+01	5.940
## factor(ctylabel)Jordan	3.223e+02	5.549e+01	5.810
## factor(ctylabel)Kenya	3.787e+02	6.301e+01	6.009
## factor(ctylabel)Korea	3.764e+02	6.504e+01	5.788
## factor(ctylabel)Kuwait	2.954e+02	5.450e+01	5.421
## factor(ctylabel)Latvia	3.147e+02	5.460e+01	5.764
## factor(ctylabel)Lesotho	3.204e+02	5.412e+01	5.919
## factor(ctylabel)Lithuania	3.246e+02	5.620e+01	5.777
## factor(ctylabel)Madagascar	3.452e+02	6.062e+01	5.694
## factor(ctylabel)Malawi	3.481e+02	5.931e+01	5.868
## factor(ctylabel)Malaysia	3.587e+02	6.168e+01	5.816
## factor(ctylabel)Mali	3.542e+02	5.950e+01	5.953
## factor(ctylabel)Mauritania	3.146e+02	5.425e+01	5.800
## factor(ctylabel)Mauritius	3.250e+02	5.173e+01	6.282
## factor(ctylabel)Mexico	3.882e+02	6.733e+01	5.766
## factor(ctylabel)Mongolia	3.166e+02	5.466e+01	5.792
## factor(ctylabel)Morocco	3.770e+02	6.322e+01	5.964

## factor(ctylabel)Mozambique	3.594e+02	6.129e+01	5.864
## factor(ctylabel)Namibia	3.231e+02	5.318e+01	6.076
## factor(ctylabel)Nepal	3.620e+02	6.204e+01	5.835
## factor(ctylabel)Nicaragua	3.279e+02	5.648e+01	5.806
## factor(ctylabel)Niger	3.535e+02	5.967e+01	5.925
## factor(ctylabel)Nigeria	4.079e+02	6.821e+01	5.980
## factor(ctylabel)Oman	3.035e+02	5.455e+01	5.564
## factor(ctylabel)Pakistan	4.438e+02	6.846e+01	6.483
## factor(ctylabel)Panama	3.220e+02	5.491e+01	5.865
## factor(ctylabel)PapuaNewGuinea	3.392e+02	5.669e+01	5.983
## factor(ctylabel)Paraguay	3.281e+02	5.656e+01	5.800
## factor(ctylabel)Peru	3.777e+02	6.237e+01	6.057
## factor(ctylabel)Philippines	3.965e+02	6.635e+01	5.976
## factor(ctylabel)Poland	3.775e+02	6.458e+01	5.846
## factor(ctylabel)Romania	3.648e+02	6.280e+01	5.808
## factor(ctylabel)Russia	4.019e+02	6.923e+01	5.806
## factor(ctylabel)Rwanda	3.580e+02	5.893e+01	6.075
## factor(ctylabel)SaudiArabia	3.470e+02	6.194e+01	5.602
## factor(ctylabel)Senegal	3.338e+02	5.842e+01	5.715
## factor(ctylabel)SierraLeone	3.365e+02	5.680e+01	5.923
## factor(ctylabel)Singapore	3.076e+02	5.578e+01	5.515
## factor(ctylabel)Slovenia	3.177e+02	5.438e+01	5.842
## factor(ctylabel)SouthAfrica	3.731e+02	6.429e+01	5.804
## factor(ctylabel)SriLanka	3.698e+02	6.151e+01	6.012
## factor(ctylabel)Swaziland	2.969e+02	5.211e+01	5.697
## factor(ctylabel)Syria	3.430e+02	6.069e+01	5.651
## factor(ctylabel)Tanzania	3.767e+02	6.362e+01	5.922
## factor(ctylabel)Thailand	4.037e+02	6.595e+01	6.122
## factor(ctylabel)Togo	3.258e+02	5.640e+01	5.777
## factor(ctylabel)Trinidad&Tobago	3.094e+02	5.217e+01	5.931
## factor(ctylabel)Tunisia	3.505e+02	5.903e+01	5.938
## factor(ctylabel)Turkey	3.921e+02	6.587e+01	5.952
## factor(ctylabel)Uganda	3.573e+02	6.200e+01	5.763
## factor(ctylabel)Ukraine	3.799e+02	6.551e+01	5.799
## factor(ctylabel)Uruguay	3.372e+02	5.573e+01	6.051
## factor(ctylabel)Venezuela	3.732e+02	6.226e+01	5.995
## factor(ctylabel)Zambia	3.491e+02	5.903e+01	5.914
## factor(ctylabel)Zimbabwe	3.444e+02	5.982e+01	5.757
##	Pr(> t)		
## l1polity	1.10e-06	***	
## l1polity_squared	0.00501	**	
## l1signed	0.48518		
## l1office	0.00560	**	
## l1gdp_pc	0.89924		
## l1lnpop	1.78e-08	***	
## l1ecris2	0.16242		
## l1bpc1	0.00832	**	
## l1avnewtar	5.41e-07	***	
## factor(ctylabel)Albania	9.58e-09	***	
## factor(ctylabel)Algeria	6.46e-09	***	
## factor(ctylabel)Argentina	5.27e-09	***	
## factor(ctylabel)Bahrain	6.24e-08	***	
## factor(ctylabel)Bangladesh	1.83e-10	***	
## factor(ctylabel)Belarus	1.82e-08	***	

## factor(ctylabel)Benin	1.92e-09	***
## factor(ctylabel)Bolivia	7.07e-09	***
## factor(ctylabel)Botswana	1.69e-09	***
## factor(ctylabel)Brazil	2.16e-09	***
## factor(ctylabel)Bulgaria	4.52e-09	***
## factor(ctylabel)Burundi	3.07e-09	***
## factor(ctylabel)Cambodia	1.96e-09	***
## factor(ctylabel)Cameroon	6.26e-09	***
## factor(ctylabel)CentralAfricanRepublic	5.71e-09	***
## factor(ctylabel)Chad	1.07e-08	***
## factor(ctylabel)Chile	8.52e-09	***
## factor(ctylabel)China	4.61e-09	***
## factor(ctylabel)Colombia	3.47e-09	***
## factor(ctylabel)Congo	5.38e-09	***
## factor(ctylabel)CostaRica	5.58e-09	***
## factor(ctylabel)Coted'Ivoire	5.95e-09	***
## factor(ctylabel)Cyprus	6.97e-09	***
## factor(ctylabel)DominicanRepublic	5.61e-09	***
## factor(ctylabel)Ecuador	3.71e-09	***
## factor(ctylabel)Egypt	2.30e-09	***
## factor(ctylabel)ElSalvador	8.09e-09	***
## factor(ctylabel)Estonia	2.28e-08	***
## factor(ctylabel)Ethiopia	7.85e-09	***
## factor(ctylabel)Fiji	9.38e-09	***
## factor(ctylabel)Gabon	4.25e-09	***
## factor(ctylabel)Gambia	1.77e-08	***
## factor(ctylabel)Ghana	1.06e-08	***
## factor(ctylabel)Guatemala	9.85e-09	***
## factor(ctylabel)Guinea	1.35e-08	***
## factor(ctylabel)Guyana	8.39e-09	***
## factor(ctylabel)Haiti	8.64e-09	***
## factor(ctylabel)Honduras	1.70e-08	***
## factor(ctylabel)Hungary	7.26e-09	***
## factor(ctylabel)India	2.22e-10	***
## factor(ctylabel)Indonesia	7.97e-09	***
## factor(ctylabel)Israel	1.42e-08	***
## factor(ctylabel)Jamaica	4.91e-09	***
## factor(ctylabel)Jordan	1.03e-08	***
## factor(ctylabel)Kenya	3.30e-09	***
## factor(ctylabel)Korea	1.17e-08	***
## factor(ctylabel)Kuwait	8.71e-08	***
## factor(ctylabel)Latvia	1.34e-08	***
## factor(ctylabel)Lesotho	5.54e-09	***
## factor(ctylabel)Lithuania	1.24e-08	***
## factor(ctylabel)Madagascar	1.97e-08	***
## factor(ctylabel)Malawi	7.42e-09	***
## factor(ctylabel)Malaysia	9.98e-09	***
## factor(ctylabel)Mali	4.56e-09	***
## factor(ctylabel)Mauritania	1.09e-08	***
## factor(ctylabel)Mauritius	6.56e-10	***
## factor(ctylabel)Mexico	1.32e-08	***
## factor(ctylabel)Mongolia	1.14e-08	***
## factor(ctylabel)Morocco	4.29e-09	***
## factor(ctylabel)Mozambique	7.60e-09	***

```

## factor(ctylabel)Namibia          2.24e-09 ***
## factor(ctylabel)Nepal            8.94e-09 ***
## factor(ctylabel)Nicaragua        1.06e-08 ***
## factor(ctylabel)Niger            5.35e-09 ***
## factor(ctylabel)Nigeria          3.91e-09 ***
## factor(ctylabel)Oman             4.03e-08 ***
## factor(ctylabel)Pakistan         1.93e-10 ***
## factor(ctylabel)Panama           7.56e-09 ***
## factor(ctylabel)PapuaNewGuinea   3.84e-09 ***
## factor(ctylabel)Paraguay         1.09e-08 ***
## factor(ctylabel)Peru             2.50e-09 ***
## factor(ctylabel)Philippines      3.99e-09 ***
## factor(ctylabel)Poland           8.39e-09 ***
## factor(ctylabel)Romania          1.04e-08 ***
## factor(ctylabel)Russia           1.06e-08 ***
## factor(ctylabel)Rwanda           2.25e-09 ***
## factor(ctylabel)SaudiArabia      3.27e-08 ***
## factor(ctylabel)Senegal          1.76e-08 ***
## factor(ctylabel)SierraLeone      5.41e-09 ***
## factor(ctylabel)Singapore       5.26e-08 ***
## factor(ctylabel)Slovenia         8.60e-09 ***
## factor(ctylabel)SouthAfrica      1.07e-08 ***
## factor(ctylabel)SriLanka         3.25e-09 ***
## factor(ctylabel)Swaziland        1.94e-08 ***
## factor(ctylabel)Syria            2.50e-08 ***
## factor(ctylabel)Tanzania         5.45e-09 ***
## factor(ctylabel)Thailand         1.71e-09 ***
## factor(ctylabel)Togo            1.24e-08 ***
## factor(ctylabel)Trinidad&Tobago  5.17e-09 ***
## factor(ctylabel)Tunisia          4.96e-09 ***
## factor(ctylabel)Turkey           4.58e-09 ***
## factor(ctylabel)Uganda           1.34e-08 ***
## factor(ctylabel)Ukraine          1.10e-08 ***
## factor(ctylabel)Uruguay          2.59e-09 ***
## factor(ctylabel)Venezuela        3.59e-09 ***
## factor(ctylabel)Zambia           5.71e-09 ***
## factor(ctylabel)Zimbabwe        1.39e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.623 on 580 degrees of freedom
## (2105 observations deleted due to missingness)
## Multiple R-squared:  0.9335, Adjusted R-squared:  0.9213
## F-statistic: 76.79 on 106 and 580 DF, p-value: < 2.2e-16

```

Interestingly, there appears to be a curvilinear relationship between the two variables! How would we interpret the results of the linear regression? How would we plot this curvilinear relationship?

```

nd <- data.frame(l1polity = seq(-10, 10, by = 1), l1polity_squared = seq(-10,
  10, by = 1)^2, l1signed = rep(0.1511, 21), l1office = rep(8.431, 21), l1gdp_pc = rep(2888,
  21), l1lnpop = rep(15.1, 21), l1ecris2 = rep(0.0641, 21), l1bpc1 = rep(0.5909,
  21), l1avnewtar = rep(14.91, 21), ctylabel = rep("Algeria", 21))

pred.p1 <- predict(main, type = "response", se.fit = TRUE, newdata = nd)

```

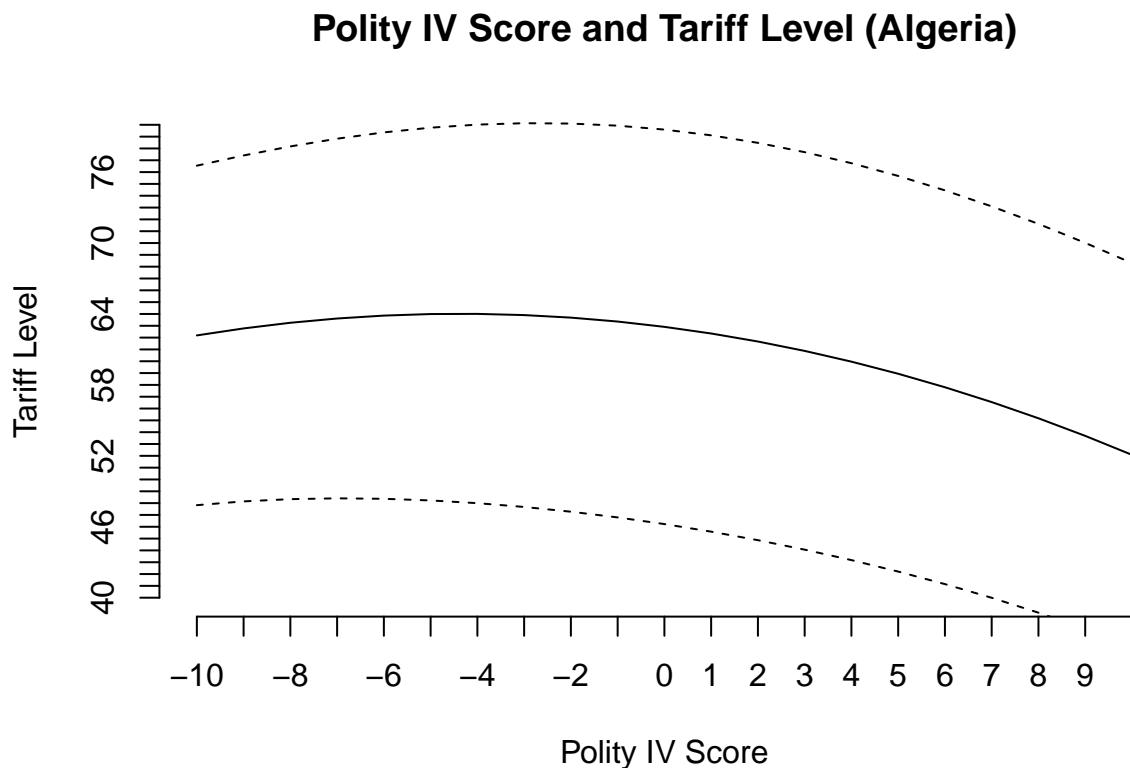
```

pred.table <- cbind(pred.p1$fit, pred.p1$se.fit)

fit <- pred.p1$fit
low <- pred.p1$fit - 2 * pred.p1$se.fit
high <- pred.p1$fit + 2 * pred.p1$se.fit
cis <- cbind(fit, low, high)

plot(pred.p1$fit, type = "l", ylim = c(40, 80), main = "Polity IV Score and Tariff Level (Algeria)",
      xlab = "Polity IV Score", ylab = "Tariff Level", axes = FALSE)
axis(1, at = seq(-10, 10, 2), labels = seq(-10, 10, 2))
axis(2, at = seq(40, 80, 4), labels = seq(40, 80, 4))
matlines(cis[, c(2, 3)], lty = 2, col = "black")

```

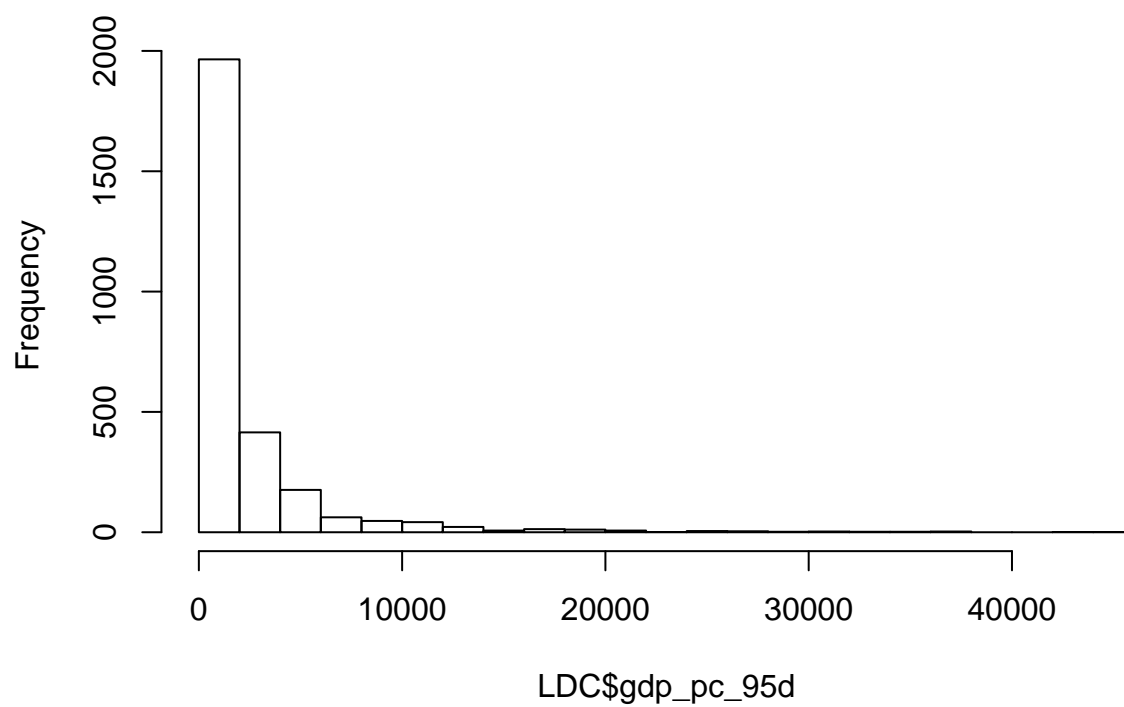


This relationship looks slightly different than the relationship estimated by Milner and Kubota.

Moreover, another frequently used way to transform data is to take the natural logarithm. In many cases, researchers do this because the distribution of the data is skewed to the right. The goal is to reduce the skewness.

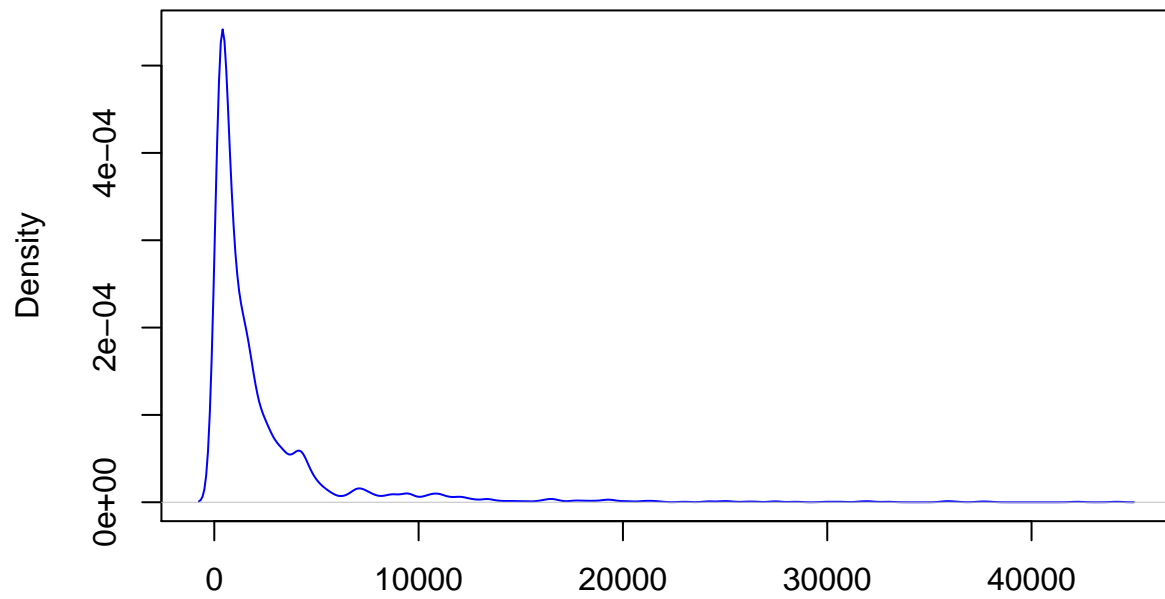
```
hist(LDC$gdp_pc_95d,breaks=21)
```


Histogram of LDC\$gdp_pc_95d



```
dens = density(LDC$gdp_pc_95d, na.rm=TRUE)
plot(dens,col="blue")
```

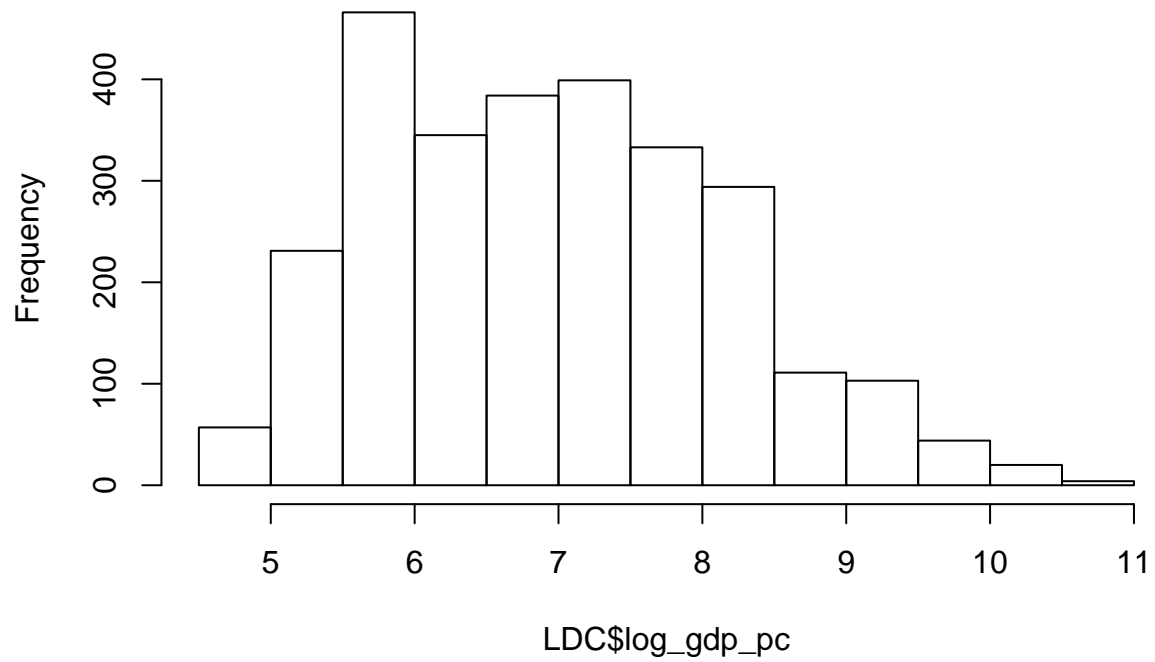
density.default(x = LDC\$gdp_pc_95d, na.rm = TRUE)



N = 2791 Bandwidth = 282.5

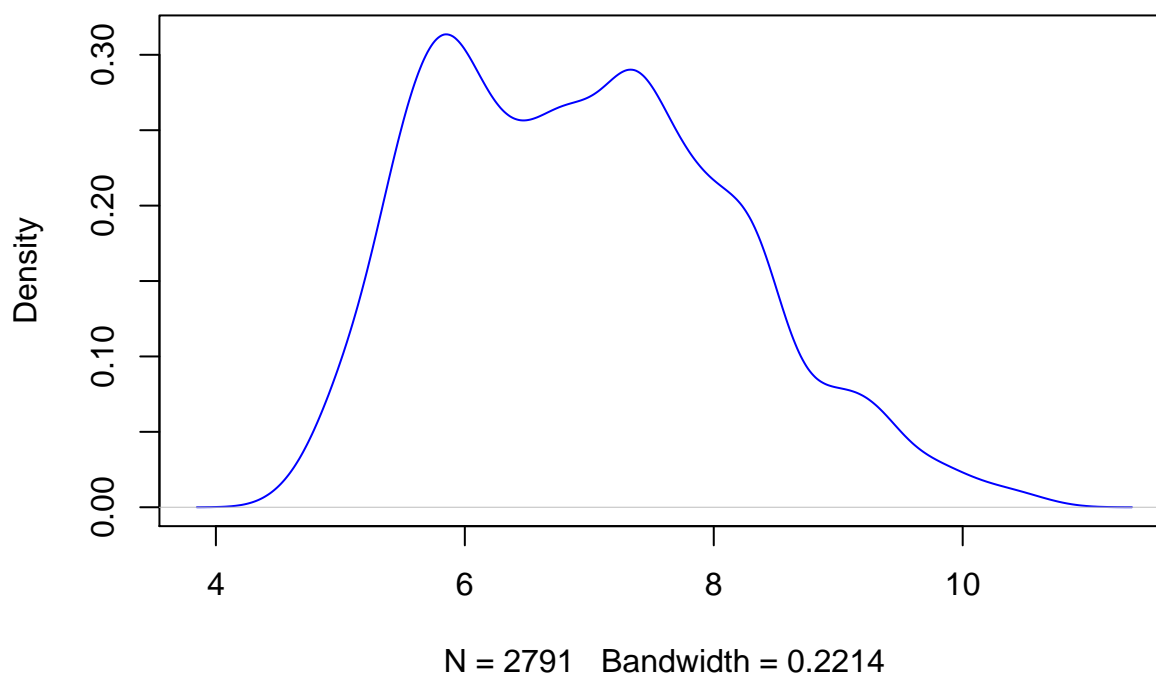
```
LDC$log_gdp_pc=log(LDC$gdp_pc_95d)
hist(LDC$log_gdp_pc,breaks=21)
```

Histogram of LDC\$log_gdp_pc



```
dens = density(LDC$log_gdp_pc, na.rm=TRUE)
plot(dens,col="blue")
```

```
density.default(x = LDC$log_gdp_pc, na.rm = TRUE)
```



```
### We have reduced the skewness and enforced a distribution that is closer to a unimodal distribution
```

3. Data management III: creating new variables

Often we can create new variables from existing ones. Let us create three categories for a country's wealth. (1) Low income countries, (2) middle income countries, and (3) high income countries.

```
LDC$incomelevel=NA
LDC$incomelevel[LDC$gdp_pc_95d<=10000]="Low-income country"
unique(LDC$incomelevel)
```

```
## [1] "Low-income country" NA
```

```
LDC$incomelevel[LDC$gdp_pc_95d > 10000 & LDC$gdp_pc_95d <= 20000] = "Middle-income country"
unique(LDC$incomelevel)
```

```
## [1] "Low-income country" "Middle-income country" NA
```

```
LDC$incomelevel[LDC$gdp_pc_95d > 20000] = "High-income country"
unique(LDC$incomelevel)
```

```
## [1] "Low-income country" "Middle-income country" "High-income country"
```

```
hist(LDC$incomelevel)
```

```
## Error in hist.default(LDC$incomelevel): 'x' must be numeric
```

```
### Does not work because it is not numeric!
```

If we want a histogram of our data, our argument needs to be numeric. We can do a simple data transformation to turn the variable into a numeric variable.

```
LDC$incomelevel[LDC$incomelevel=="Low-income country"]=0  
LDC$incomelevel[LDC$incomelevel=="Middle-income country"]=1  
LDC$incomelevel[LDC$incomelevel=="High-income country"]=2
```

```
hist(LDC$incomelevel)
```

```
## Error in hist.default(LDC$incomelevel): 'x' must be numeric
```

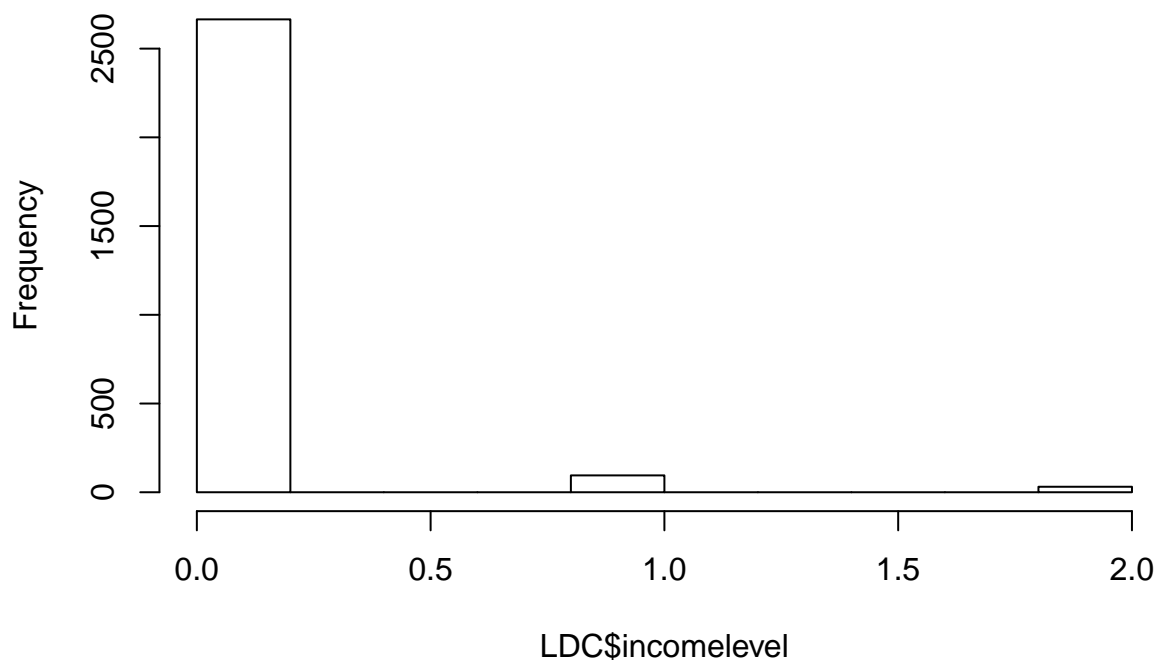
```
### Why does this not work? We just assigned numeric values.
```

We just assigned numeric values to our variable but R still treats this a character variable. What is the reason for this?

In order to change the coding of the variable to numeric, we can use the following command:

```
LDC$incomelevel=as.numeric(LDC$incomelevel)  
hist(LDC$incomelevel)
```

Histogram of LDC\$incomelevel



As we can see, the vast majority of our sample are low-income countries.

Let's say you want to create a variable that shows you if the Polity IV Score changed in any given year, with 0 indicating no change and 1 indicating a change. How would we do that?

```
LDC$politychange=NA
for (i in 2:length(LDC$polityiv_update2)){
  if (LDC$ctylabel[i] == LDC$ctylabel[i-1]){
    if (LDC$polityiv_update2[i] != LDC$polityiv_update2[i-1]){
      LDC$politychange[i]=1
    } else {
      LDC$politychange[i]=0
    }
  }
}
```

4. Data management IV: useful commands

```
unique(LDC$ctylabel)
```

```
##      [1] "Turkey"                "SouthAfrica"
##      [3] "Argentina"              "Bolivia"
##      [5] "Brazil"                 "Chile"
##      [7] "Colombia"               "CostaRica"
##      [9] "DominicanRepublic"     "Ecuador"
##     [11] "ElSalvador"            "Guatemala"
##     [13] "Haiti"                 "Honduras"
##     [15] "Mexico"                "Nicaragua"
##     [17] "Panama"                "Paraguay"
##     [19] "Peru"                  "Uruguay"
##     [21] "Venezuela"             "Guyana"
##     [23] "Jamaica"               "Trinidad&Tobago"
##     [25] "Bahrain"               "Cyprus"
##     [27] "Iran"                  "Israel"
##     [29] "Jordan"                "Kuwait"
##     [31] "Oman"                  "SaudiArabia"
##     [33] "Syria"                 "UnitedArabEmirates"
##     [35] "Egypt"                 "Bangladesh"
##     [37] "Bhutan"                "Cambodia"
##     [39] "SriLanka"              "India"
##     [41] "Indonesia"             "Korea"
##     [43] "Laos"                  "Malaysia"
##     [45] "Nepal"                 "Pakistan"
##     [47] "Philippines"           "Singapore"
##     [49] "Thailand"              "Djibouti"
##     [51] "Algeria"               "Angola"
##     [53] "Botswana"              "Burundi"
##     [55] "Cameroon"              "CentralAfricanRepublic"
##     [57] "Chad"                  "Comoros"
##     [59] "Congo"                 "Zaire"
##     [61] "Benin"                 "EquatorialGuinea"
```

```
## [63] "Ethiopia"          "Gabon"
## [65] "Gambia"            "Ghana"
## [67] "GuineaBissau"      "Guinea"
## [69] "Coted'Ivoire"      "Kenya"
## [71] "Lesotho"           "Madagascar"
## [73] "Malawi"            "Mali"
## [75] "Mauritania"        "Mauritius"
## [77] "Morocco"           "Mozambique"
## [79] "Niger"             "Nigeria"
## [81] "Zimbabwe"          "Rwanda"
## [83] "Senegal"           "SierraLeone"
## [85] "Namibia"           "Swaziland"
## [87] "Tanzania"          "Togo"
## [89] "Tunisia"           "Uganda"
## [91] "BurkinaFaso"       "Zambia"
## [93] "Fiji"              "PapuaNewGuinea"
## [95] "Armenia"           "Azerbaijan"
## [97] "Belarus"           "Albania"
## [99] "Georgia"           "Kazakhstan"
## [101] "KyrgyzRepublic"   "Bulgaria"
## [103] "Moldova"           "Russia"
## [105] "Tajikistan"        "China"
## [107] "Turkmenistan"      "Ukraine"
## [109] "Uzbekistan"        "Estonia"
## [111] "Latvia"            "Hungary"
## [113] "Lithuania"         "Mongolia"
## [115] "Croatia"           "Slovenia"
## [117] "Poland"            "Romania"
```

```
# Displays all of the empirically observed values
```

```
head(LDC$fdignp)
```

```
## NULL
```

```
# Returns the first five values
```

```
names(LDC)
```

```
## [1] "ctylabel"          "date"              "polityiv_update2"
## [4] "gdp_pc_95d"        "newtar"            "l1polity"
## [7] "l1polity.1"        "l1signed"          "l1office"
## [10] "l1gdp_pc"          "l1lnpop"           "l1ecris2"
## [13] "l1bpc1"            "l1avnewtar"        "polityiv_squared"
## [16] "l1polity_squared"  "log_gdp_pc"        "incomelevel"
## [19] "politychange"
```

```
# Returns the variable names of a data frame
```

```
class(LDC$ctylabel)
```

```
## [1] "character"
```

```
# Show the classification of a variable

char_newtar=as.character(LDC$newtar)
# Changes a variable to a character variable

val_newtar=as.numeric(LDC$char_newtar)
# Changes a variable to a numeric variable

quantile(LDC$newtar,p=c(0.1,0.9), na.rm=T)
```

```
## 10% 90%
## 8.5 40.0
```

```
# Displays the 10th and 90th percentile of a variable
```

5. The most useful R packages for applied work

The following are some of the most useful packages for applied work. I recommend to get the related text books and download the documentations of these packages. These packages have many useful commands that can help you to deal with data management and data analysis.

1. car — associated with Fox & Weisberg’s book “Companion to Applied Regression”
2. arm — associated with Gelman & Hill’s book on “Regression and Multilevel/Hierarchical Models”
3. Zelig — by political science professor Gary King
4. ggplot2 — introduced extensively in Chang’s “R Graphics Cookbook”
5. stargazer