

dataset B summary

Peder M. Isager

11/1/2019

```
data.B <- readRDS("../processed_data/dataset_B.rds")

## summary of key variables

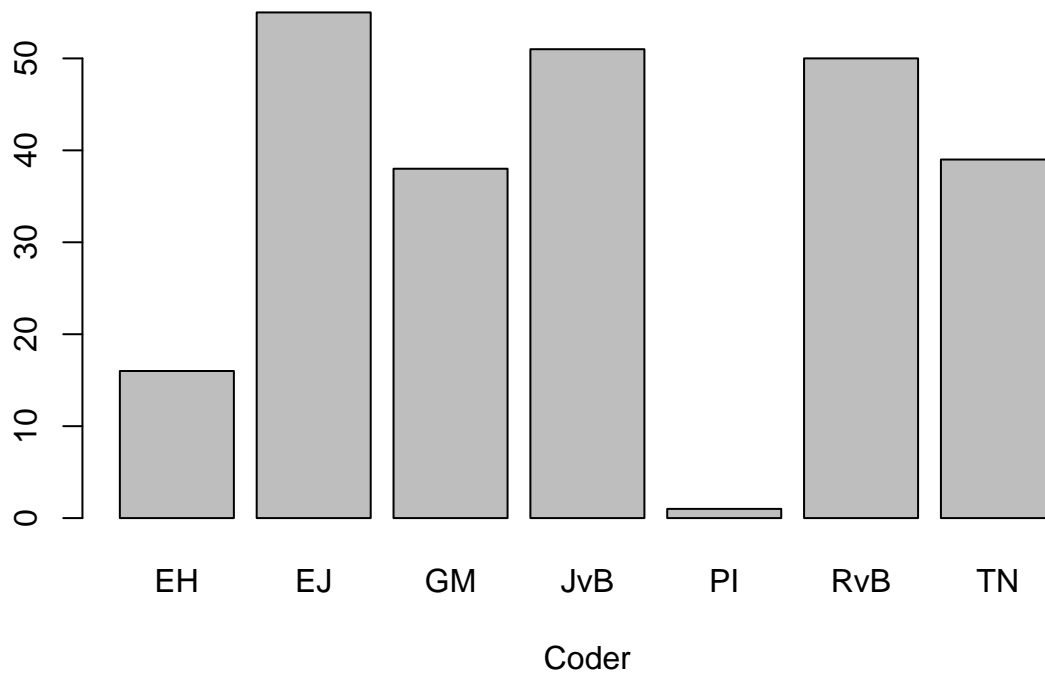
key.vars <- c("PY", "study_number", "sample_size", "coder", "resolver", "excluded", "TC", "RV")

summary(data.B[, key.vars])
```

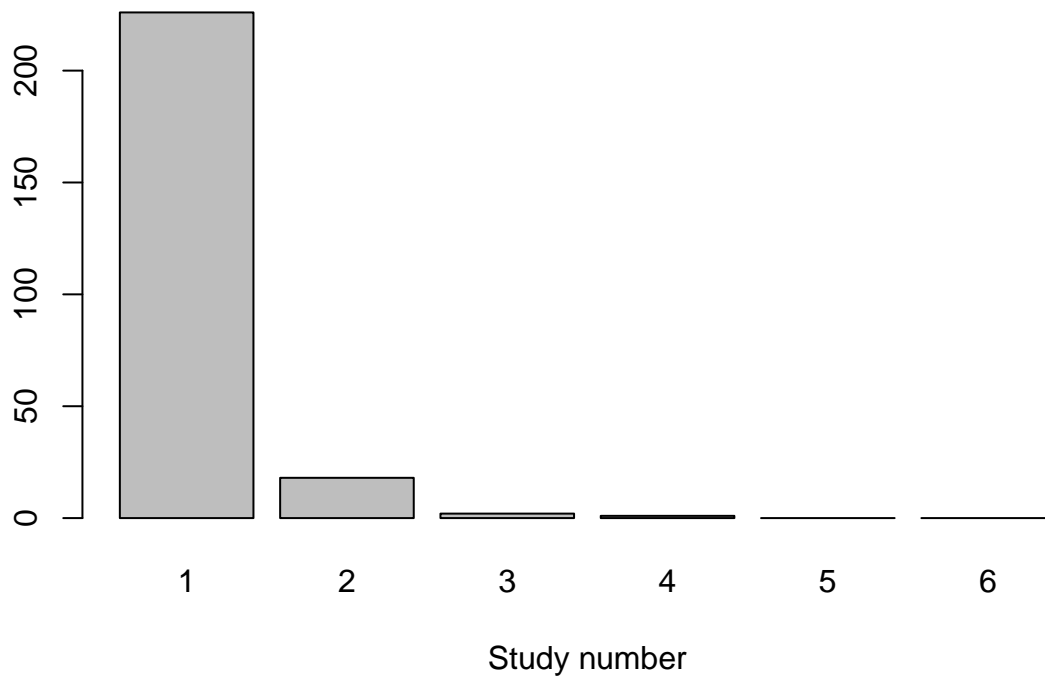
##	PY	study_number	sample_size	coder	resolver
##	Min. :2009	1 :226	Min. : 1.00	EH :16	AV : 20
##	1st Qu.:2012	2 : 18	1st Qu.: 18.00	EJ :55	PI : 34
##	Median :2014	3 : 2	Median : 23.00	GM :38	NA's:196
##	Mean :2014	4 : 1	Mean : 47.09	JvB:51	
##	3rd Qu.:2016	5 : 0	3rd Qu.: 35.00	PI : 1	
##	Max. :2019	6 : 0	Max. :1235.00	RvB:50	
##		NA's: 3		TN :39	
##	excluded	TC	RV		
##	Min. :0	Min. : 0.00	Min. :0.00000		
##	1st Qu.:0	1st Qu.: 3.00	1st Qu.:0.03191		
##	Median :0	Median : 9.00	Median :0.09122		
##	Mean :0	Mean : 18.32	Mean :0.16794		
##	3rd Qu.:0	3rd Qu.: 24.00	3rd Qu.:0.20000		
##	Max. :0	Max. :134.00	Max. :1.71429		
##			NA's :5		

```
## visualization of key variables

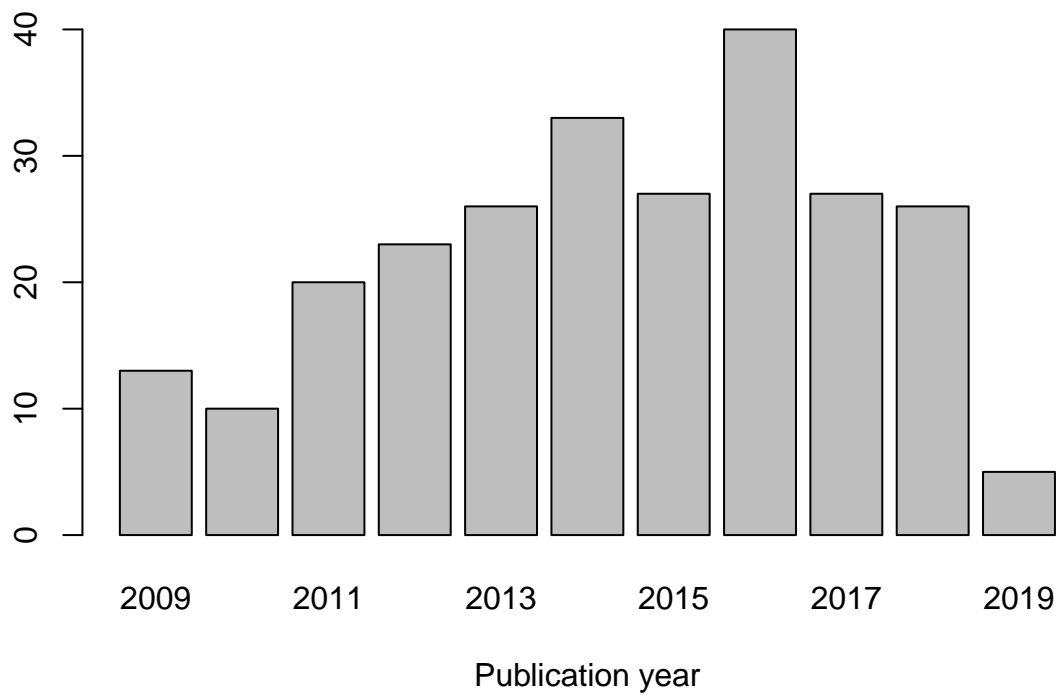
coder.freq <- table(data.B$coder)
barplot(coder.freq, xlab = "Coder") # Plot frequency of coders
```



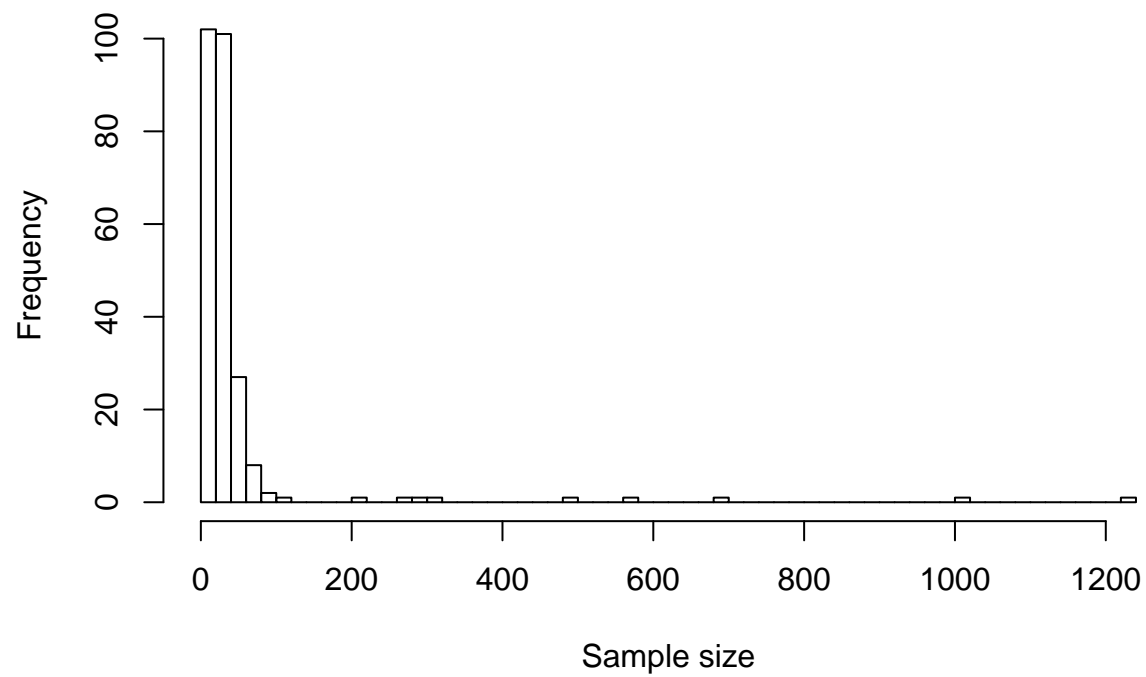
```
studyn.freq <- table(data.B$study_number)
barplot(studyn.freq, xlab = "Study number") # Plot frequency of study numbers (first vs. second vs. fo
```



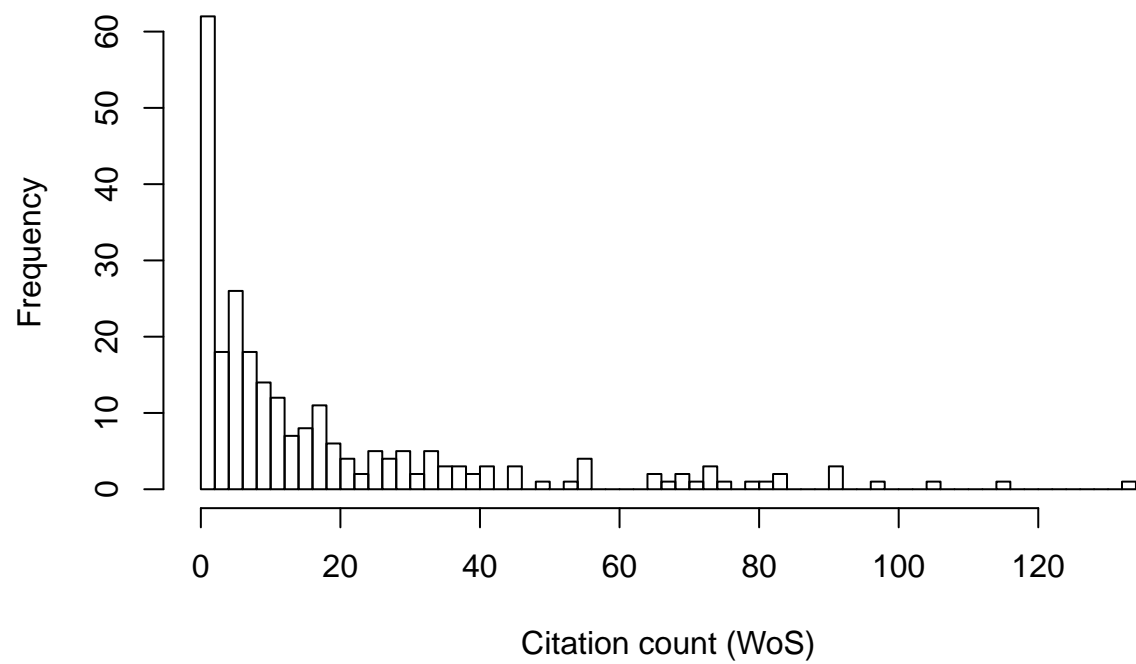
```
pubyear.freq <- table(data.B$PY)
barplot(pubyear.freq, xlab = "Publication year") # Plot frequency of publication years
```



```
hist(data.B$sample_size, breaks = 50, xlab = "Sample size", main = "") # Plot sample size distribution
```



```
hist(data.B$TC, breaks = 50, xlab = "Citation count (WoS)", main = "") # Plot citation count distribut
```



```
hist(data.B$RV, breaks = 50, xlab = "Replication value (C/Y)/(N-3)", main = "") # Plot replication val
```

