Homework 3

Anna Ziff

R Workflow for Economists

1. Using the `words` vector in the `stringr` library, solve the following once with one regular expression and once with multiple uses of `str_subset()`.[1]

   a. What are the words that start with x or end with x?
   b. Find all words that start with a vowel and end with a consonant.

2. Download the 2014 Master Plan from Singapore's data portal. Find the shapefile in your download and load it into R. Download the data about residents by subzone from the same portal. Load it into R. How many subzones are there in each dataset? Merge the two datasets only keeping observations that appear in both datasets. Inspect the resulting data carefully, paying close attention to the interpretation of each row.

3. Using `tmap` create a map showing Singapore and its subzones. Second, create a map in which the subzones are colored in according to the number of residents between ages 45 and 49 in 2001. Facet the map by sex. Add whatever formatting elements you would like.

4. The below code uses the library `boot` to calculate the bootstrapped standard errors of the regression relating the log number of subscriptions to journals and their log price per citations. Write the analogous code using one of the functions for parallel computing in chapter 8. What are some pros and cons of using the `boot` package versus coding your own (parallel) bootstrap?

```
library(AER)
library(boot)
library(tidyverse)
```

```
# Load data
data("Journals")
journals <- Journals %>%
    mutate(citeprice = price/citations) %>%
    tibble()

# Point estimates
out <- lm(log(subs) ~ log(citeprice), data = journals)

# Function for boot()
refit <- function(data, i) {
    lm(log(subs) ~ log(citeprice), data = data[i, ]) %>%
        coefficients()
}

# Implement bootstrap function with 1000 replications
out_boot <- boot(journals, refit, R = 1000)
```

5. Read chapters 1-5 (they are very short chapters) of Hadley Wickham's style guide. During this class,

[1] This question comes from this chapter, to which you may want to refer.

I have broken some of these suggestions. For your own work, you should follow the conventions that lead to the best code possible, balancing efficiency, generality, and readability. “Fix up” the below code following the advice from Wickham’s book.

```r
mean=matrix( rnorm(100), nrow=10);

for(row_iterator in 1:10) { for (col_iterator in 1:10) {
if(mean[   row_iterator ,col_iterator] < 0){ mean[row_iterator,col_iterator] <- 0 }}}

# Define a character matrix
CHARvec=rep(c("a", 'b',"c", 'Hello "world"'), 25);
# Define a logical vector
BooleanVector<-rep(c(T, F, T, T), 25)
```

6. Optional: Complete a regex crossword.

7. Optional: Read about measuring performance and complete the exercises. Beyond this particular chapter, this textbook is a great resource to gain a more advanced understanding of R.