

Assignment 2

Anh Le

October 1, 2014

```
## knitr configuration: http://yihui.name/knitr/options#chunk_options
opts_chunk$set(comment = "", error= TRUE, warning = FALSE, message = FALSE,
               tidy = FALSE, cache = F, echo = T,
               fig.width = 4, fig.height = 4, fig.align="center")
```

1 Question 1

Let's pursue further the in-class example of ordering / selecting one variable based on another. We have the following mock data.

```
mock_data <- data.frame(country=c("US", "UK", "South Africa", "Liberia"),
                        region=c("America", "Europe", "Africa", "Africa"),
                        gdppc=c(40000, 35000, 25000, 9000),
                        stringsAsFactors=FALSE)
```

```
mock_data
```

	country	region	gdppc
1	US	America	40000
2	UK	Europe	35000
3	South Africa	Africa	25000
4	Liberia	Africa	9000

What I showed you in class is to select the country variable based on the gdppc variable, like so:

```
# Get countries with GDP per capita > 10000
mock_data$country[mock_data$gdppc > 10000]

[1] "US"          "UK"          "South Africa"

# Get countries with above-average GDP per capita
mock_data$country[mock_data$gdppc > mean(mock_data$gdppc)]

[1] "US" "UK"
```

Now, the question is how to select countries that have `gdppc > 10000` AND belong in Africa? Phrased more generally, how do we subset the data frame using two / multiple conditions? (Google if you don't know how – I have phrased the question in very Google-able terms)

So here's your first assignment.

1. Using the mock data above, select African countries that have `gdppc > 10000`.
2. Download real data from package `WDI`, then subset the data according to some conditions that interests you. (E.g. List all African countries that have below / above average GDP per capita; What about other continents? Variables other than GDP, etc.)

2 Question 2 – Problem 1.9.1 in the book

This problem involves data frame – re-read the book chapter on data frame if necessary

```
library(DAAG) # install if you have it yet
```

The following table gives the size of the floor area (ha) and the price (\$000), for 15 houses sold in the Canberra (Australia) suburb of Aranda in 1999.

```
houseprices
  area bedrooms sale.price
9   694         4    192.0
10  905         4    215.0
11  802         4    215.0
12 1366         4    274.0
13  716         4    112.7
14  963         4    185.0
15  821         4    212.0
16  714         4    220.0
17 1018         4    276.0
18  887         4    260.0
19  790         4    221.5
20  696         5    255.0
21  771         5    260.0
22 1006         5    293.0
23 1191         6    375.0
```

1. Plot `sale.price` versus `area`.
2. Use the `hist()` command to plot a histogram of the sale prices.
3. Repeat (a) and (b) after taking logarithms of sale prices.

4. The two histograms emphasize different parts of the range of sale prices. Describe the differences.