Assignment 1 - Solution

Anh Le

October 3, 2014

Use package `WDI` (an R package that downloads data from the World Bank's World Development Indicators) , download data and plot it. Do not download data by hand, which is both inefficient and not reproducible. Your grader should be able to run your code from start to finish and get the same result.

Here are some ideas to help you start. Feel free to do things you think are cool.

- Easy: Plot the histogram of countries' GDP / your favorite statistic in one year

- Easy: Plot the GDP of one country across years

- Harder: Plot the GDP of multiple countries / continents / groups across years

- Harder: Add text label, color, etc. to highlight certain countries

Make sure to have appropriate title, axis labels, etc.

P/S: There are other packages similar to `WDI` to download Political Science data straight into R. Amazing time that we live in.

```r
library("WDI") # install.packages as necessary

## Loading required package:  RJSONIO

help(WDI) # Read the docs -- this one is particularly comprehensible
```

1   Solution

```r
## knitr configuration: http://yihui.name/knitr/options#chunk_options
opts_chunk$set(comment = "", error= TRUE, warning = FALSE, message = FALSE,
               tidy = TRUE, cache = F, echo = T,
               fig.width = 5, fig.height = 5, fig.align="center")
```

```r
rm(list = ls())  # Always do this to start with a clean slate
library("WDI")
library("ggplot2")
library("zoo")  # To deal with time, year, etc.
```

1.1 US military expenditure

```r
search_res <- WDIsearch(string = "military")
US_military <- WDI(country = "US", indicator = c("MS.MIL.XPND.CN", "MS.MIL.XPND.GD.ZS"),
    start = 1970, end = 2010)
colnames(US_military)[c(4, 5)] <- c("mil_exp", "mil_exp_gdp")  # Rename columns
US_military <- US_military[complete.cases(US_military), ]  # Remove missing data

# Convert column year to Date data type so that R understands that it's a
# date instead of a number
US_military$year <- as.Date(strptime(US_military$year, format = "%Y"))

plot(US_military$year, US_military$mil_exp, type = "l", xlab = "Year", ylab = "Military expe
    main = "US military expenditure across years")
abline(v = as.Date("2001", "%Y"), col = "red")
text(x = as.Date("2003", "%Y"), y = quantile(US_military$mil_exp, probs = 0.9),
    "Year 2001")
```

US military expenditure across years
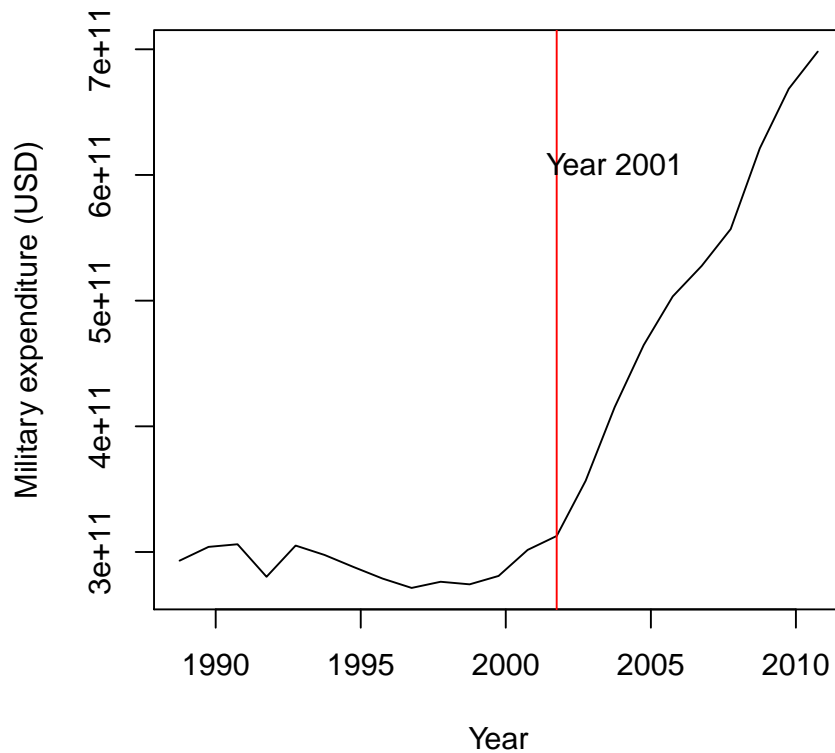
1.2 World GDP by regions (ggplot2 demonstration)

```r
rm(list = ls())
world_gdp <- WDI(country = "all", indicator = c("NY.GDP.PCAP.CD"), extra = TRUE,
    start = 1970, end = 2010)

# Remove aggregate regions (i.e. OECD, etc.)
world_gdp <- world_gdp[world_gdp$region != "Aggregates", ]

# Remove cases with missing data
world_gdp <- world_gdp[complete.cases(world_gdp), ]

# as.yearmon is a function in the zoo package, converting string to datetime
# p is just a common name for a plot object
p <- ggplot(data = world_gdp, aes(x = as.Date(as.yearmon(year)), y = NY.GDP.PCAP.CD)) +
```
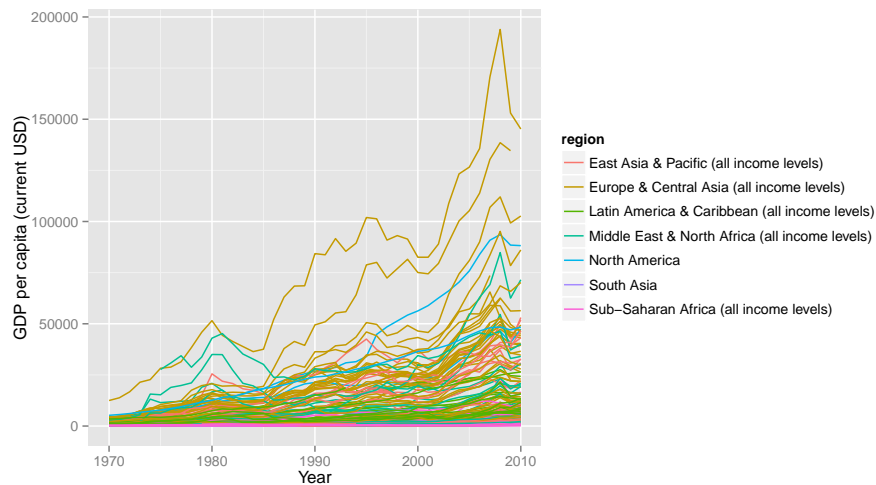
```
    labs(x = "Year", y = "GDP per capita (current USD)")

p_all_countries <- p + geom_line(aes(group = country, color = region))
p_all_countries
```


Hmm, despite all the talk about China and East Asian miracles, the countries who saw the strongest growth is in Europe & Central Asia. I wonder what those countries are?

```
# Select only the rows where the year is 2009
world_gdp_2009 <- world_gdp[world_gdp$year == 2009, ]

# Select only the rows where its GDP per cap is in the top 5 percent
world_gdp_2009_top_5_percent <- world_gdp_2009[world_gdp_2009$NY.GDP.PCAP.CD >=
    quantile(world_gdp_2009$NY.GDP.PCAP.CD, 0.95), ]

# geom_text() will add text to the plot. In this case, the text label is
# country name
p_rich_countries <- p_all_countries + geom_text(data = world_gdp_2009_top_5_percent,
    aes(x = as.Date("2009", "%Y"), y = NY.GDP.PCAP.CD, label = country))
p_rich_countries
```
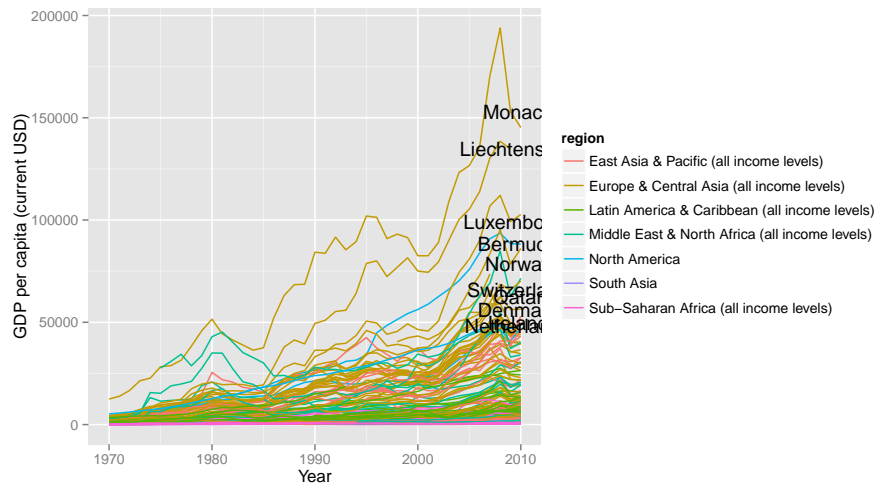
Let's see where are the East Asian miracles in the plot. US for scale.

```
world_gdp_select <- world_gdp[world_gdp$country %in% c("China", "Japan", "Korea, Rep.",
    "United States"), ]

# I blur the countries' line with alpha = 0.25 Then I add geom_line() and
# geom_text() for the selected countries
p + geom_line(aes(group = country, color = region), alpha = 0.25) + geom_line(data = world_g
    aes(group = country), size = 1) + geom_text(data = world_gdp_select[world_gdp_select$yea
    2010, ], aes(x = as.Date("2009", "%Y"), y = NY.GDP.PCAP.CD, label = country))
```