《国际关系定量分析基础》2020 秋季

第二次小组作业 (共计 100 分)

成员名字 1　　成员名字 2　　成员名字 3　　成员名字 4　　成员名字 5

截止时间：**2020 年 11 月 2 日 11：59 am**

面板数据（panel data）也称为横截面时间序列数据（cross-sectional time-series data），是国际关系中最常用的一种数据格式。本次作业的目标是练习如何利用既有数据库建立一个整洁的包含世界所有国家 1990-2016 年的面板数据，并添加国家特征相关的其他数据和变量。既有数据包括世界银行（world bank）、Pen World Table 以及 terrorism 数据库等。

注意事项：

- 小组作业截止时间：**2020 年 11 月 2 日 11：59 am**
- 请将文件解压缩后，直接在 R Markdown 文件中完成本次作业
- 作业在网络学堂提交，每个小组仅需提交一份
- 提交作业的文件名需以 `HW-2-Team-X.Rmd`，`HW-2-Team-X.pdf` 或者 `HW-2-Team-X.html`, 请将 `X` 替换为小组编号，如 `HW-2-Team-A.Rmd`、`HW-2-Team-A.pdf` 或 `HW-2-Team-A.html`。(若 R Markdown 出现无法 `knit` 为 `pdf` 情况，则使用 `bookdown::html_document2:` 会生成为 `html`)
- 请显示每道题的 R Code 于 pdf 中，注重 Code 的整洁性和可读性，可参考Google's R Style Guide
- 本次作业所需 R Packages 已经提供。本次作业需要的数据已经提供，请将数据与 `HW-2-Team-X.Rmd` 放在同一工作路径的文件夹内，通过 `load("terrorism.RData")` 加载。

```
load("terrorism.RData")
load("wdi.RData")
```

创建基本数据框（共 20 分）

1.(**20 分**) 请用 R 中的 `states`、`countrycode` 和 `dplyr` 三个 packages, 创建一个 1990-2016 年包括全世界所有国家的，以年为单位的面板数据，并将数据框命名为 `base_df`。**注意**：创建数据应使

用基于 COW 的国家代码 (country code)，数据应该删除台湾省（台湾的 COW code 是 713），最终的数据只保留 `cowcode`, `country_name` 和 `year` 三个变量（参考表-1 的输出结果）。结合 `base_df` 数据，请回答这个数据的“分析单元”（unit-of-analysis）是什么?

```
# 请完成代码
library(states)
library(countrycode)
library(dplyr)
base_df <- state_panel()
```

处理 WDI 数据（共 20 分）

世界银行发展指标（world bank development indicators）是社会科学研究中常用的关于国家层次的政治经济和社会数据。其中，R 中的 `WDI` 是一个基于世界银行数据的软件包，记录了世界银行及其开发指标的相关数据。利用 `WDI` 软件可包获取 1990-2016 年所有国家的包括“军费开支占 GDP 比重”、“GDP 经济增长率”、“FDI 占 GDP 比重的数据”，其对应的变量分布为 `MS.MIL.XPND.GD.ZS`, `NY.GDP.MKTP.KD.ZG`, `BX.KLT.DINV.WD.GD.ZS`。通过如下代码或者 `load("wdi.RData")` 可以获得这一数据。

```
# Foreign direct investment, net inflows (% of GDP): BX.KLT.DINV.WD.GD.ZS
# Military expenditure (% of GDP): MS.MIL.XPND.GD.ZS
# GDP growth (annual %): NY.GDP.MKTP.KD.ZG
library(WDI)
wdi <- WDI(country = "all", indicator = c("MS.MIL.XPND.GD.ZS",
                                          "NY.GDP.MKTP.KD.ZG",
                                          "BX.KLT.DINV.WD.GD.ZS"),
           start = 1990, end = 2016)
```

2.(**10 分**) 请利用 `dplyr` 软件包，将数据 `wdi` 中以上三个变量重命名为 `mil_expen` (MS.MIL.XPND.GD.ZS), `gdp_growth` (NY.GDP.MKTP.KD.ZG), `fdi_percent` (BX.KLT.DINV.WD.GD.ZS)。同时，利用 R 中的 `countrycode` 包，创建一个名为 `ccode` 变量——该变量应当使用 COW 的国家编码。最后，基于 `wdi` 数据，只保留如下变量：`country`, `ccode`, `year`, `mil_expen`, `gdp_growth`, `fdi_percent` (参考表-2 的输出结果)。

```
load("wdi.RData")
wdi <- wdi %>%
      rename()
```

3.(**10 分**) 基于 `wdi` 数据，利用 `dplyr` 包，以表格形式显示 2015 年全世界军费开支占 GDP 比重最高的 10 个国家（参考表-3 的输出结果）。**提示**：可利用 `dplyr` 包中 `filter`, `slice_max`、`arrange` 三个命令。

```
# 请完成代码
wdi %>%
    dplyr::filter()
```

处理 Penn Word Table 9.1 数据（共 20 分）

4.(**10 分**) Penn Word Table 是经济学家常用的关于世界各国经济指标的数据，R 中的 `pwt9` 软件包已经收录了 1950-2017 年各国的相关数据。利用 `pwt9` 包，获取 1990-2016 年之间各国的数据，将数据框命名为 `pwt9`；同时利用 COW 的国家编码，创建一个 `ccode` 变量。**提示**：数据应该是 4450 行，48 列。

```
# 请完成代码
library(pwt9)
data("pwt9.1")
pwt9 <- pwt9.0 %>%
      filter()
```

5.(**10 分**) 基于创建的 1990-2016 年的 `pwt9` 数据框，只保留各国 1990-2016 年间的 GDP 变量（对应的变量名为 `rgdpna`）、人口 (对应的变量名为 `pop`)。**提示**: 最后保留的 `ccode`, `year`, `pop`, `rgdpna` 变量和观察量如表-4 所示.

```
# 请完成代码
pwt9 <- pwt9 %>%
        filter() %>%
        select()
```

处理 Terrorism 数据（共 20 分）

6.(**10 分**) `globalterrorism.RData` 记录了 1990-2016 年发生在世界各国的恐怖袭击事件。首先，描述这一数据的分析单位（unit-of-analysis）是什么? 其次，根据 `globalterrorism.RData` 这一数据，利用 `dplyr` 包，将数据汇总到国家--年层次，以显示各国在每一年发生的恐怖袭击的次数和每年的伤亡人数总和。**提示**：需要使用 `group_by`, `summarise` 等命令。

```
# 请完成代码
load("globalterrorism.RData")
gtd <- globalterrorism %>%
       mutate() %>%
       group_by()
```

合并三个数据（共 10 分）

7.(**10 分**) 分别将 wdi, pwt9 以及 gtd 合并到 base_df 中，产生一个包含 11 个变量，5094 个观测量的数据。注意需要将 sum_nkillter 与 sum_events 中的 NA 重新赋值为 0. **提示**：使用 left_join 这一命令（参考如下最终结果）。

```
# 请完成代码
base_df <- left_join()
```

可视化数据（10）

8.(**5 分**) 利用 stargazer 包，基于合并的 base_df 数据，制作变量间的描述统计表，产生关于如下变量的描述性统计表格（见表-6 所示）。**提示**: 可以通过 help(stargazer) 选择对应的统计量(summary.stat)。

```
# 请完成代码
library(stargazer)
```

9.(**5 分**) 利用 GGally 包，基于合并的 base_df 数据，绘制包括如下变量的（图-1）相关系数图。

```
# 请完成代码
library(GGally)
```