

《国际关系定量分析基础》2020 秋季

第一次小组作业 (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 10 月 12 日 11: 59 am

东南亚地区 (如图 1) 是国际关系和比较政治学界关注的重点地区，本次作业将利用公开数据，对东南亚地区国家的政治、经济、社会、外交等关系进行描述。

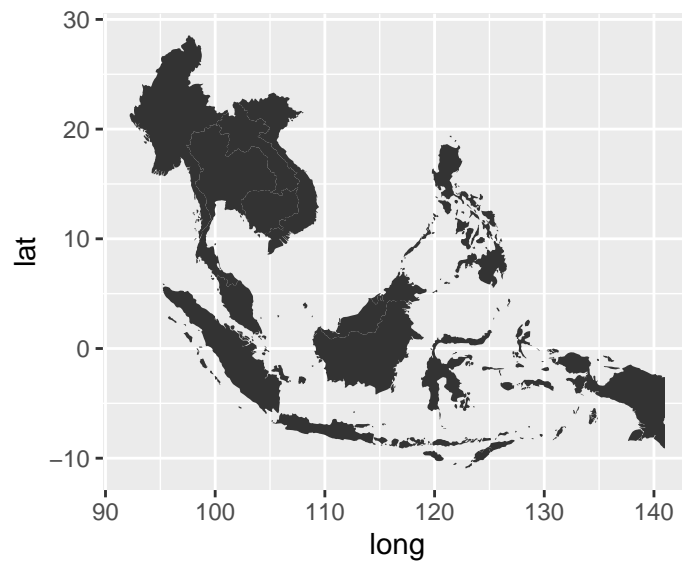


图 1: 东南亚地图

注意事项：

- 小组作业截止时间：2020 年 10 月 12 日 11: 59 am
- 请直接在 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-1-Team-X.Rmd 和 HW-1-Team-X.pdf(或者 HW-1-Team-X.html), 请将 X 替换为小组编号，如 HW-1-Team-A.Rmd 和 HW-1-Team-A.pdf。(若 R Markdown 出现无法 knit 为 pdf 情况，可以使用 bookdown::html_document2:，则会生成 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
- 本次作业所需的数据和 R Packages 已经提供。本次作业需要的数据可以通过以下命令获取（或直接 `load("terrorism.RData")`）:

```
load(url("https://cc458.github.io/files/terrorism.RData"))
load(url("https://cc458.github.io/files/conflict.RData"))
```

其中，

- `terrorism.RData` 包括三个数据集: `region_map_shp`, `gtd_region`, `wdi`;
- `conflict.RData` 包括四个数据集 `polity`, `acled_cnty`, `ideal_point_wide`, `ucdp_cnty`

R 基础问题（共 15 分）

1.(10 分) 请用 `knitr::kable` 创建一个 8*3 的表格，总结这 7 个数据框。表格除第一行为 header 外，其余每一行表示一个数据框；除第一列为数据框的名称外，其余两列分别为每一个数据框的变量 (variables) 和观测量 (observations) 数。（提示：可以先创建一个包含这些信息的新数据框，然后再使用 `kable` 创建表格；也可以利用 R Markdown 手动创建）。

```
library(knitr)
# 请将代码写在此处；获取数据维度的命令为 dim()
dim(wid)
# 创建一个数据新的数据集
df <- data.frame(dataset = c(),
                  variable_num = c(),
                  observations_num = c())
# 利用 kable 产生一个新的表格
kable(df, caption = " 数据集信息")
```

2.(5 分)`stargazer` 是政治学常用的产生统计表格的软件包，请利用 `stargazer` 提供一个关于 `wdi` 数据的描述性统计表格。

```
# 加载软件包
library(stargazer)
# 使用 stargazer 命令产生新的描述性表格
stargazer(, header=FALSE, type='latex', title = " 描述性统计结果",
          digit.separator = "")
```

数据可视化问题（共 85 分）

3.(10) 数据 `region_map_shp` 是一个包含空间信息的数据集，其中变量 `terratck` 包含了东南亚 1991-2006 年之间所遭遇的恐怖主义袭击数量总和。同时数据集 `gtd_region` 记录了东南亚 1991-2006 年每一次恐怖袭击的经纬度地理位置（变量 `longitude` 和 `latitude`）。请利用 `ggplot2` 这一软件包产生如下地图（图 2），描述各国在此期间的恐怖主义数量分布以及被袭击的地点，并简要描述你对关于东南亚恐怖袭击活动地理分布的观察。

此题的部分代码如下

```
ggplot() +  
  geom_polygon(data = ,  
               aes(x = , y = , group = group, fill = ),  
               size = 0.25) +  
  coord_fixed() +  
  scale_fill_manual(values = c("#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7"),  
                    name = "Total number of attacks",  
                    na.value = "gray",  
                    labels = c("1 - 10", "10 - 100", "100 - 500", "500 - 1000",  
                                ">2000", "NA")) +  
  geom_point(aes(x = , y = , color = "black"),  
             alpha=.5, size = 0.5, na.rm= FALSE) +  
  scale_color_manual(name = "",  
                     labels = c("Attack locations"),  
                     values = "black") +  
  labs(y = "latitude", x = "longitude")
```

4.(10 分) 数据集 `wdi` 中的变量 `gdpgrowth` 记录了东南亚各国在 1990-2018 年之间的国民生产总值增长率。请利用 `ggplot2` 绘制各国在此时间段内的国民生产总值增长率随时间而变化的折线图（图 3）。并据此图简要谈谈 1997-1998 年亚洲金融危机对东南亚国家经济增长的冲击（提示：可以使用 `facet_wrap` 分别绘制，也可以绘制到同一张图）

此题的部分代码如下

```
ggplot() +  
  geom_point(colour = "pink", size = 2)+  
  geom_line() +  
  labs(x = "Year", y = "GDP per capita growth (annual %)",  
       caption = "Source: World Bank Group 2020") +  
  scale_x_continuous(breaks = seq(1990, 2018, 10)) +
```

```
facet_wrap() +  
theme(strip.text.x = element_text(size = 10, color='white',angle=0),  
      strip.background = element_rect(fill = "#525252", color='#525252'))
```

5.(10 分) 数据集 wdi 中 gdpcc 表示人均国民生产总值 (GDP per capita, 以 2018 年美元为单位), 变量 milexp 表示军费开支占国民生产总值的比值。请利用 ggplot2 描述 gdpcc 与 milexp 之间的关系, 并讨论你是否发现什么规律。

请在此完成代码

```
ggplot() +
```

6.(10 分) 数据集 wdi 中 fdi 表示当年外国直接投资 (净流入) 占国民生产总值的比例。请 ggplot2 及其 facet_wrap 命令, 描述东南亚各国在 1990-2018 年之间外国直接投资的变化情况, 并据此简要讨论你观察到何种模式和规律。(提示: 参考第 4 题)

请在此完成代码

```
ggplot() +
```

7.(10 分) 请利用可视化方法, 简要描述并讨论数据集 wdi 中 fdi, gdpcc, gdpgrowth 和 milexp 这四个变量的相关关系。

请在此完成代码

8.(10 分) 数据集 polity 是国际关系中最常见的用来测量国家整体类型的数据。其中的变量 polity2 的值域为 [-10, +10], 即 “最不民主”(-10) 到 “最民主”(10)。请利用 ggplot 简要描述这一变量的分布情况。

请在此完成代码

```
ggplot() +geom_
```

9.(10 分) 数据集 acled_cnty 记录了 2010-2019 年东南亚国家经历的 “抗议” (protest) 和 “骚乱” (riots) 数量。利用 ggplot2 绘制各国经历的抗议和骚乱变化情况, 并比较抗议和骚乱在各国内部的差异情况。(提示: 利用 ggplot 中的 linetype 和 facet_wrap 命令)

此题的部分代码如下

```
ggplot(, linetype= ) +  
  geom_point()+  
  geom_line() +  
  labs(x = "Year", y = "Social unrest",  
        caption="Source: ACLED 2020") +  
  scale_x_continuous(breaks = seq(1991, 2019, 5)) +
```

```
facet_wrap() +
theme(strip.text.x = element_text(size = 12, color='white',angle=0),
      legend.position = "bottom",
      legend.title = element_blank(),
      strip.background = element_rect(fill = "#525252", color='#525252'))
```

10.(5 分) 数据集 `ucdp_cnty` 记录了东南亚国家 1946-2019 年之间经历的三种冲突(`type_of_conflict2`: 国内冲突、国际冲突以及国际化的国内冲突) 的数量 (`conflict`)。请利用 `ggplot` 描述各国分别经历的这三类冲突分布情况。注意其中的缺失值 (NA)，并简要说明为何有缺失值。

此题的部分代码如下

```
ggplot(, aes(x = , y = , fill = )) +
  geom_bar(position="dodge", stat="identity") +
  scale_fill_manual(values=c("yellow", "red", "blue")) +
  labs(x = "", y = "Count", caption="Source: UCDP 2019") +
  theme_bw() +
  theme(strip.text.x = element_text(size = 12, color='white',angle=0),
        legend.position = "bottom",
        legend.title = element_blank(),
        strip.background = element_rect(fill = "#525252", color='#525252'))
```

11.(10 分) 数据集 `ideal_point_wide` 记录了东南亚国家在 1973-2018 年之间在联合国大会中投票是否同意中国、印度、美国和俄罗斯（苏联）的情况。利用 `ggplot` 分布绘制东南亚国家在此期间的立场变化。

此题的部分代码如下

```
ggplot(, aes(), color= ) +
  geom_line(aes(color=)) +
  labs(x = "Year", y = "UNGA voting agreement",
       caption = "Source: Bailey, Strezhnev, and Voeten (2017, JCR)") +
  scale_x_continuous(breaks = seq(1973, 2018, 10)) +
  facet_wrap() +
  theme(strip.text.x = element_text(size = 12, color='white',angle=0),
        legend.position = "bottom",
        legend.title = element_blank(),
        strip.background = element_rect(fill = "#525252", color='#525252')) +
  scale_colour_manual(values = c("blue","green","gray", "red"))
```