

《国际关系定量分析基础》2020 秋季

第一次小组作业 (共计 100 分)

参考答案

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。

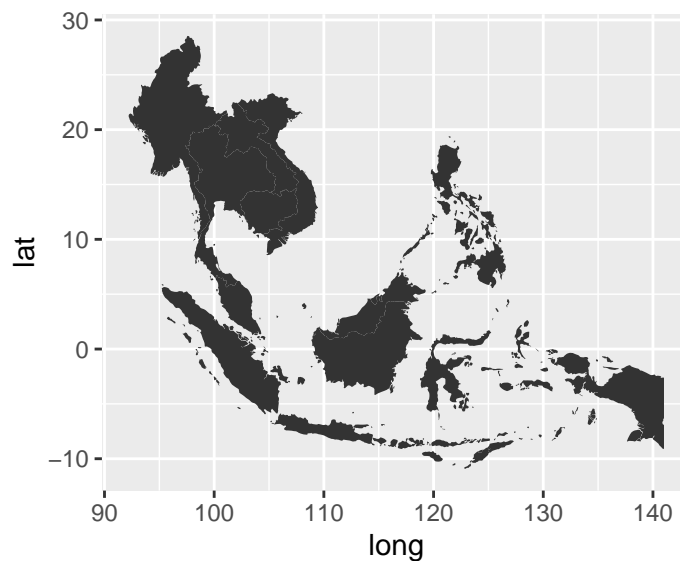


图 1: 东南亚地图

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)

df <- data.frame(dataset = c("region_map_shp", "gtd_region", "wdi", "polity",
                             "acled_cnty", "ideal_point_wide", "ucdp_cnty"),
                 obs_num = c(dim(region_map_shp)[1], dim(gtd_region)[1],
                              dim(wdi)[1], dim(polity)[1],
                              dim(acled_cnty)[1], dim(ideal_point_wide)[1],
                              dim(ucdp_cnty)[1]),
                 variable_num = c(dim(region_map_shp)[2], dim(gtd_region)[2],
                                   dim(wdi)[2], dim(polity)[2],
                                   dim(acled_cnty)[2], dim(ideal_point_wide)[2],
                                   dim(ucdp_cnty)[2]))

knitr::kable(df, caption = " 数据集信息")
```

另外一种办法是

```
data.frame(dataset = c("region_map_shp", "gtd_region", "wdi", "polity",
                        "acled_cnty", "ideal_point_wide", "ucdp_cnty")) %>%
  dplyr::bind_cols(., data.frame(t(mapply(dim, list(region_map_shp, gtd_region, wdi,
```

表 1: 数据集信息

dataset	obs_num	variable_num
region_map_shp	17769	18
gtd_region	9585	136
wdi	319	10
polity	241	4
acled_cnty	107	4
ideal_point_wide	1664	5
ucdp_cnty	17	5

表 2: 数据集信息

dataset	obs_num	variable_num
region_map_shp	17769	18
gtd_region	9585	136
wdi	319	10
polity	241	4
acled_cnty	107	4
ideal_point_wide	1664	5
ucdp_cnty	17	5

```
                                polity, acled_cnty, ideal_point_wide,
                                ucdp_cnty)))) %>%
dplyr::rename(obs_num = X1, variable_num = X2) %>%
kable(., 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。

```
library(stargazer)
stargazer(wdi, header=FALSE, type = "latex", title = " 描述性统计结果",
          digit.separator = "")
```

表 3: 描述性统计结果

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
year	319	2004.000	8.380	1990	1997	2011	2018
gdppc	296	7568.071	13061.030	95.188	715.872	6037.739	66188.780
gdpgrowth	304	3.707	4.478	-37.002	2.232	5.885	14.173
fdi	292	5.078	5.349	-2.757	1.861	6.110	28.598
gender	134	0.639	0.381	0.000	0.333	1.000	1.000
milexp	283	2.310	1.558	0.190	1.265	3.083	8.675
poverty	82	14.833	17.348	0.000	0.875	22.775	66.700
pop	319	50243727.000	63772880.000	258721	4677894.0	69610521	267663435

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。

```
ggplot() +
  geom_polygon(data = region_map_shp,
              aes(x = long, y = lat, group = group, fill = terratck),
              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(data = gtd_region, aes(x = longitude, y = latitude, 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",
      caption="Source: GTD 2019")
```

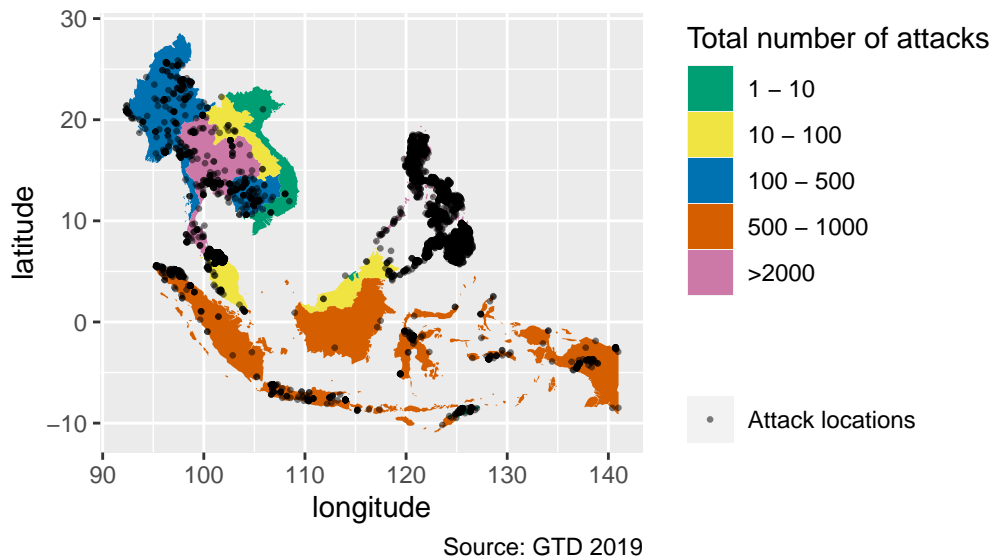
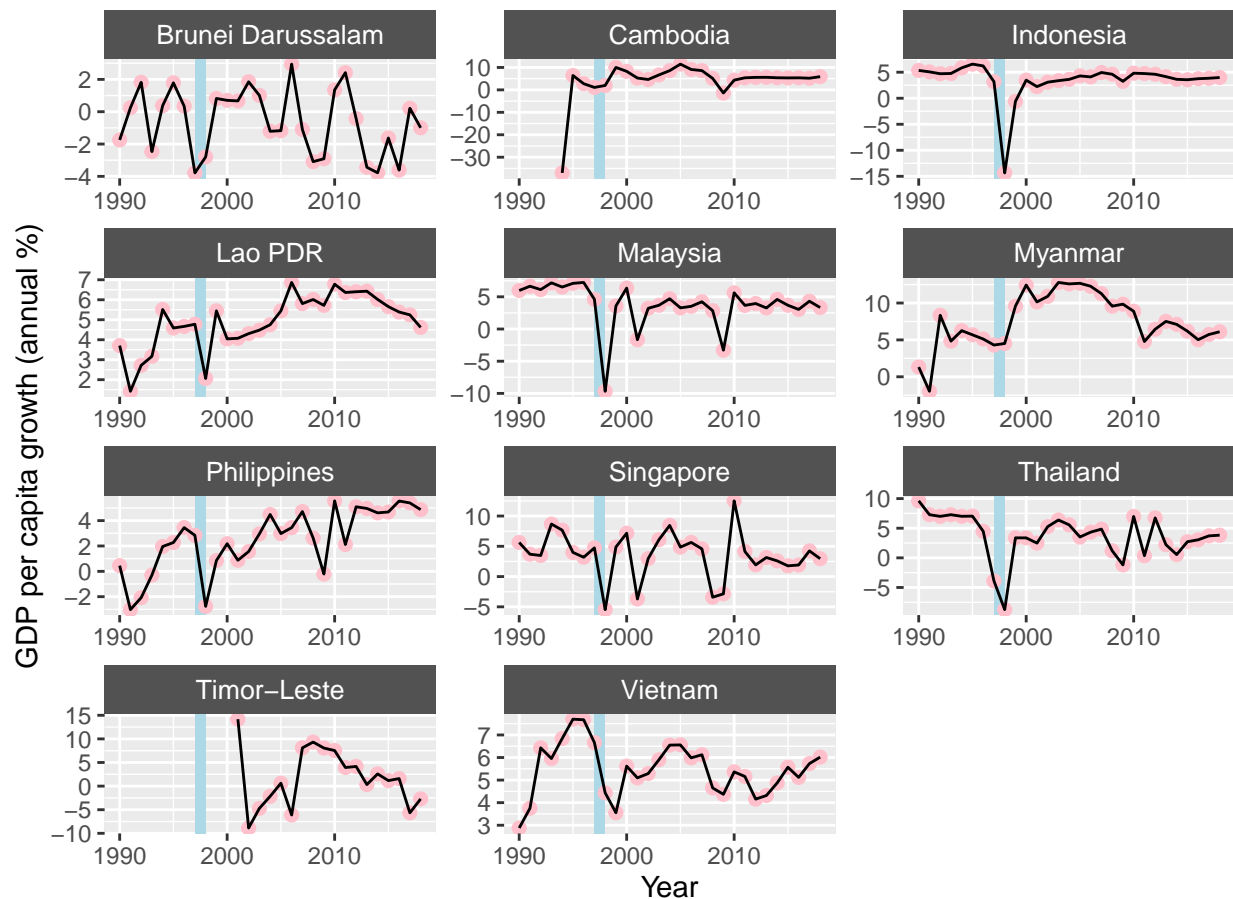


图 2: 东南亚恐怖袭击数量分布。

- 恐怖袭击活动在东南亚分布呈现出国别差异，以菲律宾和泰国最多，越南最少。

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)

```
ggplot(wdi, aes(year, gdpgrowth)) +
  #code to display interval for 1997-1998
  geom_rect(aes(xmin = 1997, xmax = 1998, ymin = -Inf, ymax = Inf), fill = "lightblue") +
  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(~country, scales = "free", nrow = 4, ncol = 3) +
  theme(strip.text.x = element_text(size = 10, color = 'white', angle = 0),
        strip.background = element_rect(fill = "#525252", color = "#525252"))
```



Source: World Bank Group 2020

图 3: 东南亚各国国民生产总值增长率变化 (1990-2018)。

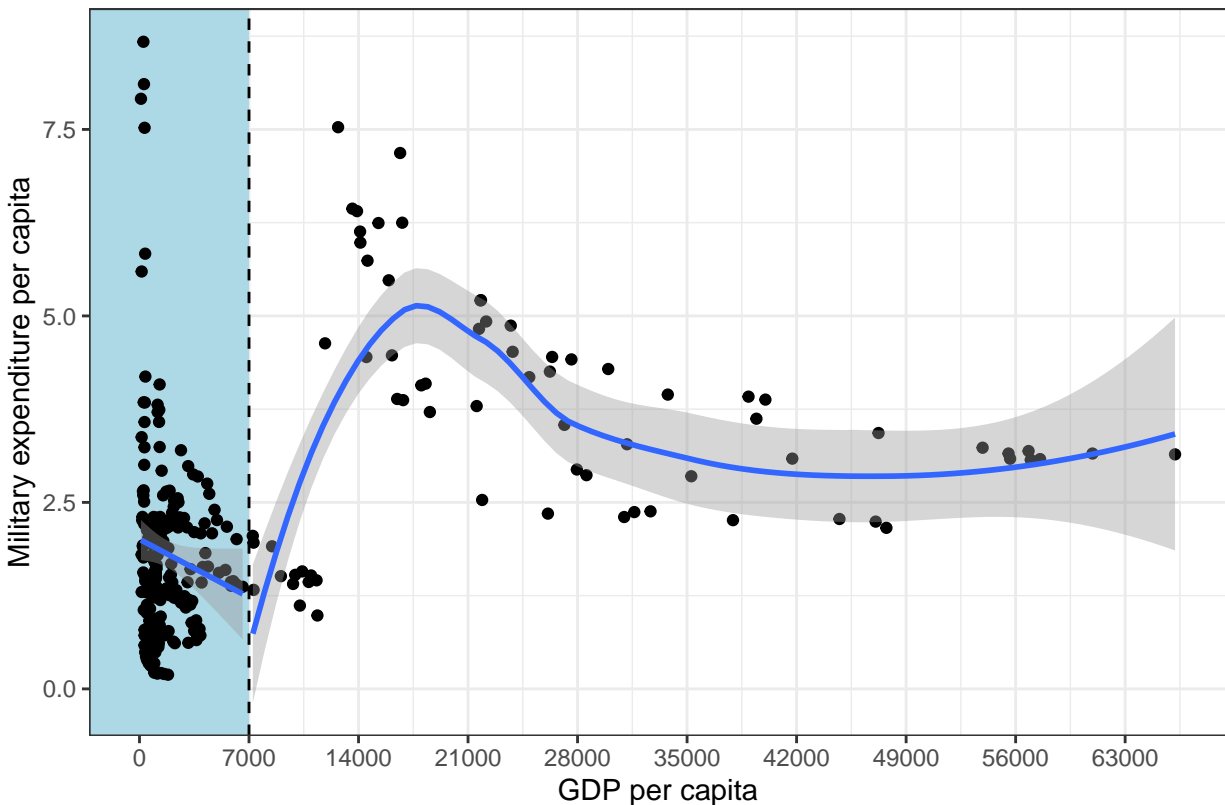
- 淡蓝色区域表示 1997-1998 年，从中可以看到东南亚各国的经济受到冲击非常严重，呈现断崖式下跌

5.(10 分) 数据集 `wdi` 中 `gdppc` 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 `gdppc` 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(data = wdi) +
  geom_rect(aes(xmin = -Inf, xmax = 7000, ymin = -Inf, ymax = Inf), fill = "lightblue") +
  geom_point(aes(x = gdppc, y = milexp)) +
  geom_smooth(data = subset(wdi, gdppc >= 7000), method = 'loess',
    aes(x = gdppc, y = milexp)) +
  geom_smooth(data = subset(wdi, gdppc < 7000), method = 'lm',
    aes(x = gdppc, y = milexp)) +
```

```
geom_vline(aes(xintercept = 7000), fill="red", linetype="dashed")+
scale_x_continuous(breaks = seq(0, 70000, by = 7000))+
labs(x = "GDP per capita",y = "Military expenditure per capita",
caption="Source: World Bank Group 2020") +
theme_bw() +
theme(strip.text.x = element_text(size = 10, color='white', angle=0),
strip.background = element_rect(fill = "#525252", color='#525252'))
```



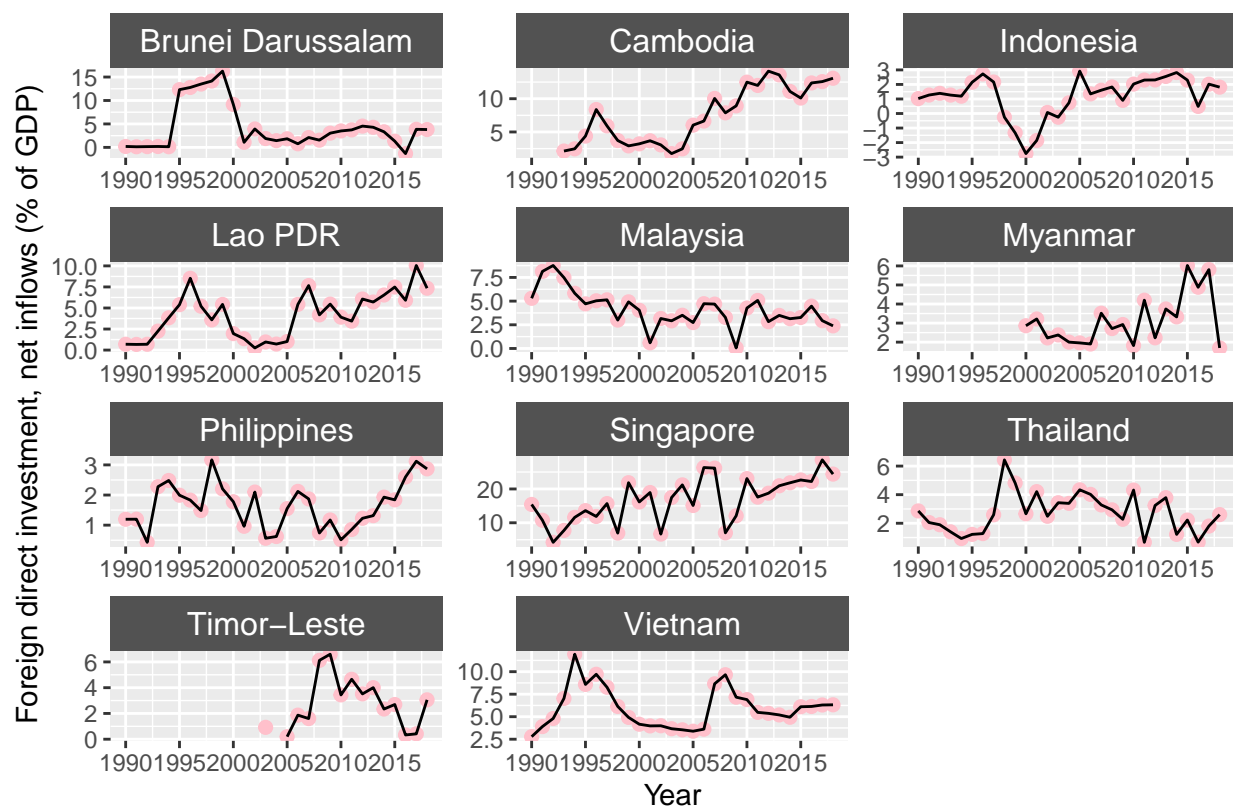
Source: World Bank Group 2020

图 4: 东南亚人均国民生产总值与军费开支关系。

- 人均国民生产总值与人均军费开支并不是简单的线性关系：在人均国民生产总值低于 7000 美元（图中淡蓝色区域）（或 10000 美元）时，二者呈现出负相关性；超过 7000 美元至 17000 美元时，呈现出正相关趋势，随后则又呈现出负相关关系。

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）

```
ggplot(wdi, aes(year, fdi)) +
  geom_point(colour = "pink", size = 2) +
  geom_line() +
  labs(x = "Year", y = "Foreign direct investment, net inflows (% of GDP)",
       caption="Source: World Bank Group 2020") +
  scale_x_continuous(breaks = seq(1990, 2018, 5)) +
  facet_wrap(~country, scales = "free", nrow = 4, ncol = 3) +
  theme(strip.text.x = element_text(size = 12, color='white',
                                     angle=0),
        strip.background = element_rect(fill = "#525252", color='#525252'))
```



Source: World Bank Group 2020

图 5: 东南亚外国直接投资。

- 各国的 FDI 在 1997-98 年的金融危机期间开始呈现出大幅度下降趋势，尤其是印度尼西亚甚至出现了负比例

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

请在此完成代码

```
library(GGally)
ggpairs(wdi[, c("fdi", "gdppc", "gdpgrowth", "milexp")],
        lower=list(continuous='smooth', combo='facetdensity'))
```

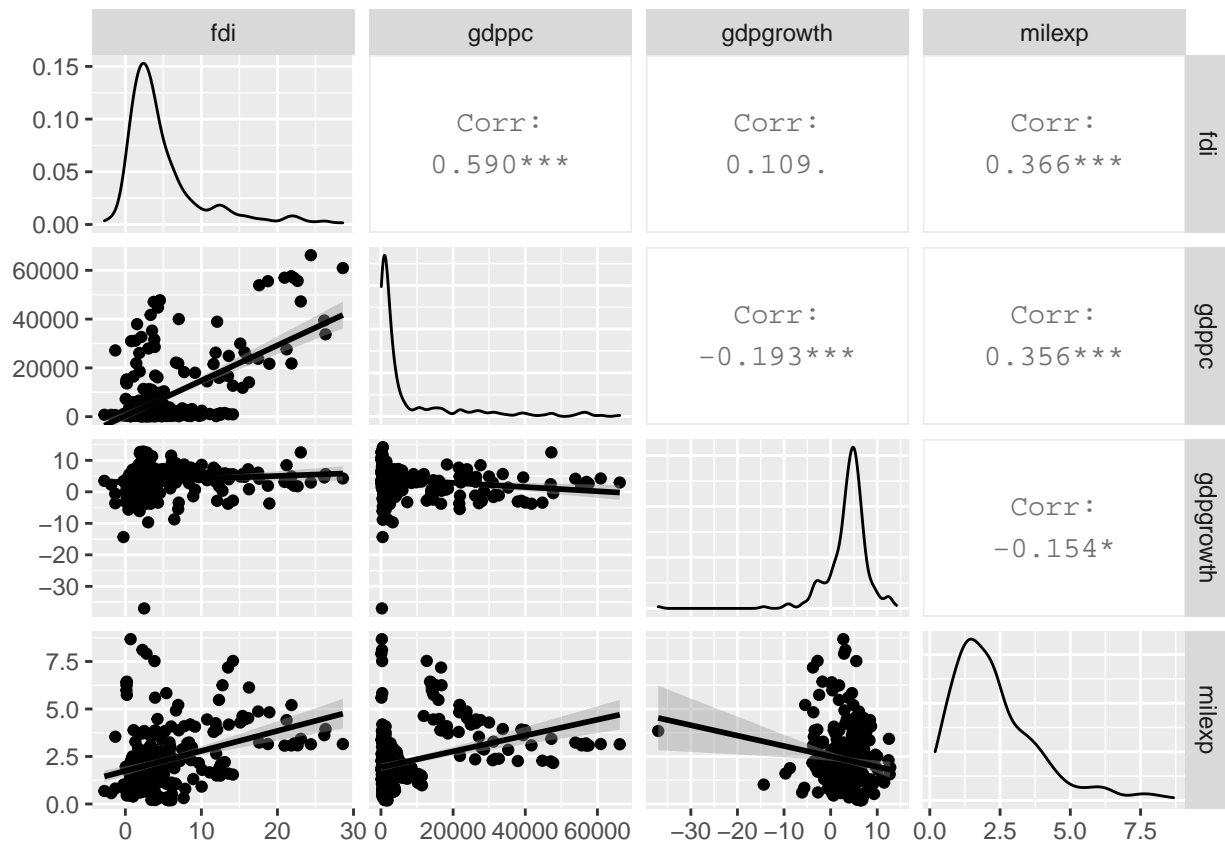


图 6: 变量间的相关关系。

- 人均国民生产总值与 FDI 之间的相关性系数为 0.59，表明二者具有高度的正相关性
- 人均军费开支与 GDP 增长率的相关性系数为-0.154，表明二者是负相关

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

合并 east timor, timor leste

```
polity <- polity %>%
```

```
  mutate(country = ifelse(country == "Timor Leste", "East Timor", country))
```

```
library(scales) #Pretty breaks for date/times
```

```
ggplot(polity, aes(y = polity2, x= year)) +
  geom_line(size = 1)+
  scale_x_continuous(breaks = seq(1991, 2018, 5)) +
  scale_y_continuous(breaks= breaks_pretty(1))+ #only one break
  labs(x = "Year", y = "Polity Score",
       caption="Source: Polity IV") +
  facet_wrap(~country, scales = "free_y", nrow = 4, ncol = 3) +
  theme(strip.text.x = element_text(size = 12, color='white',
                                     angle=0),
        strip.background = element_rect(fill = "#525252", color='#525252'))
```

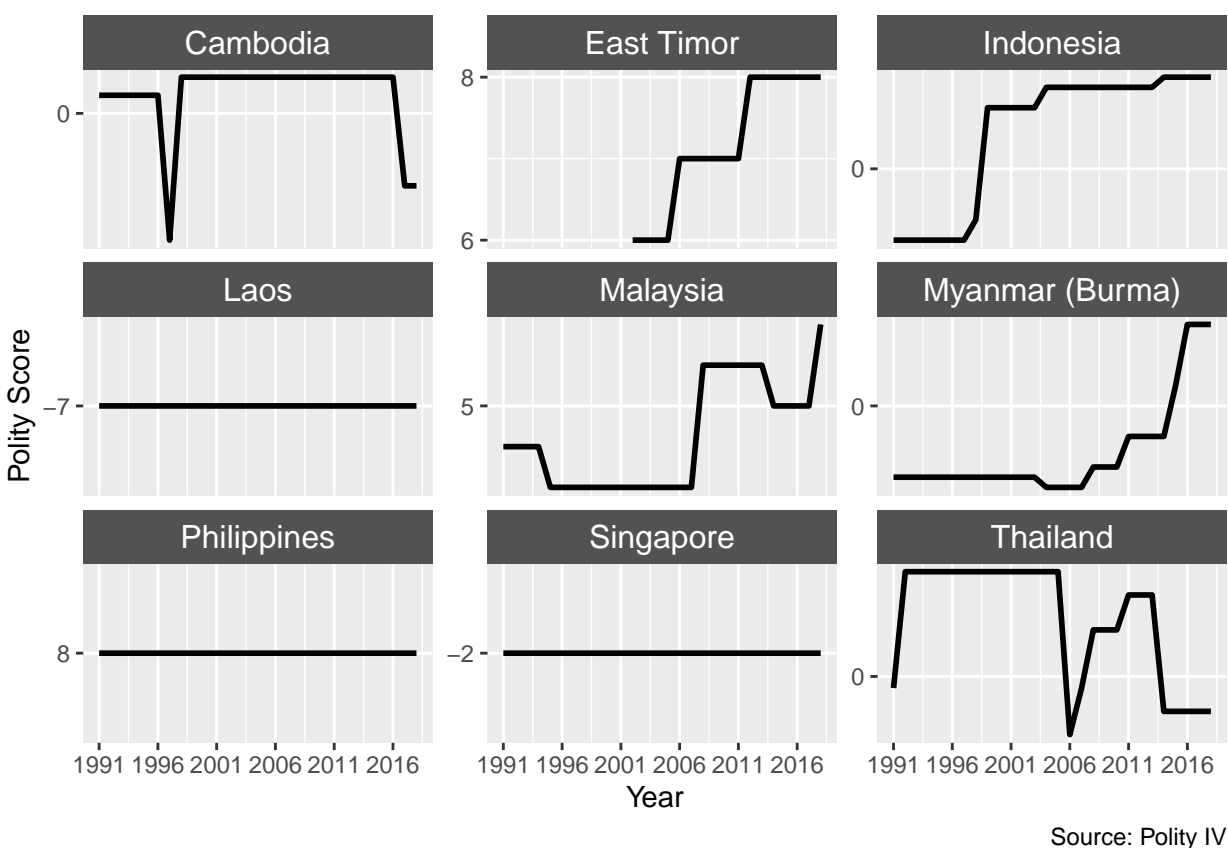
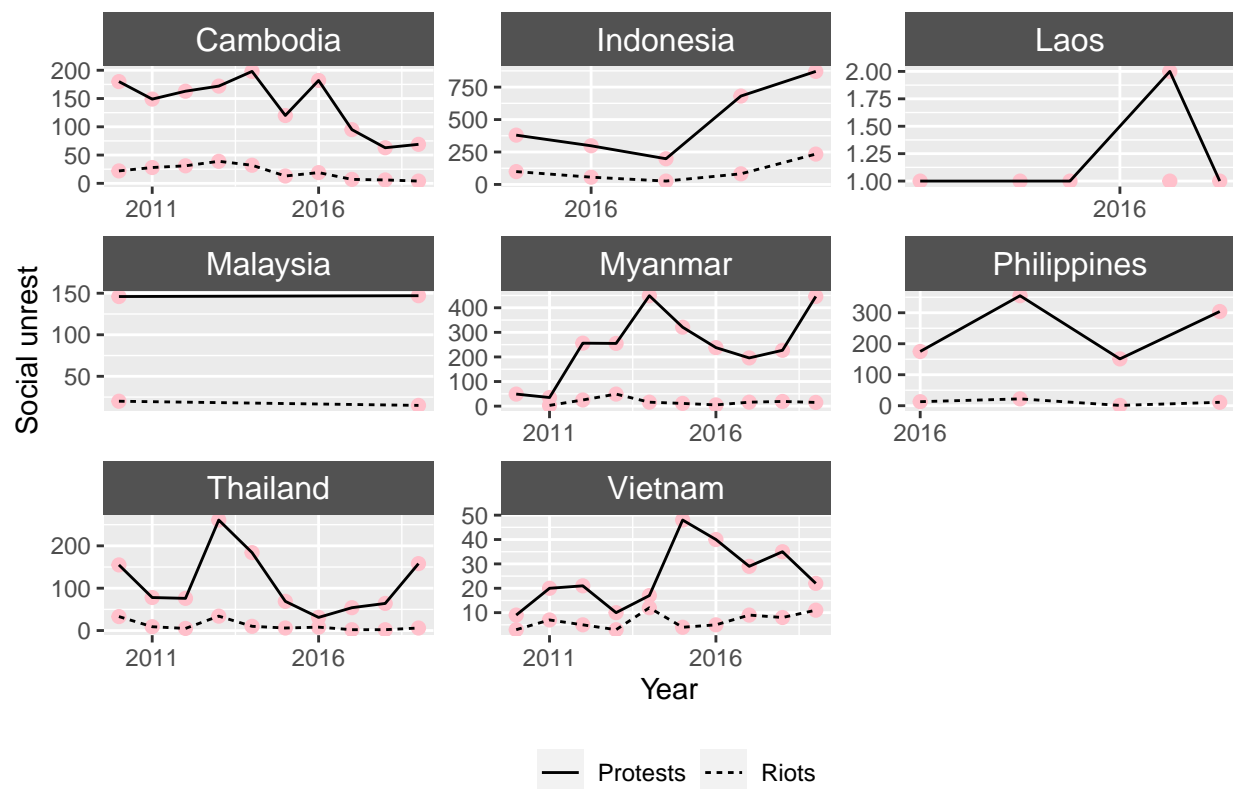


图 7: 东南亚国家政体变化。

- 泰国 2006 年经历了一政变，政体分值变动大
- 菲律宾、新加坡的政体得分比较稳定，但是菲律宾的得分高于新加坡，表明菲律宾更为民主

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)

```
ggplot(acled_cnty, aes(year, events, group = event_type),
      linetype= event_type) +
  geom_point(colour = "pink", size = 2)+
  geom_line(aes(linetype=event_type)) +
  labs(x = "Year", y = "Social unrest",
       caption="Source: ACLED 2020") +
  scale_x_continuous(breaks = seq(1991, 2019, 5)) +
  facet_wrap(~country, scales = "free", nrow = 4, ncol = 3) +
  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'))
```



Source: ACLED 2020

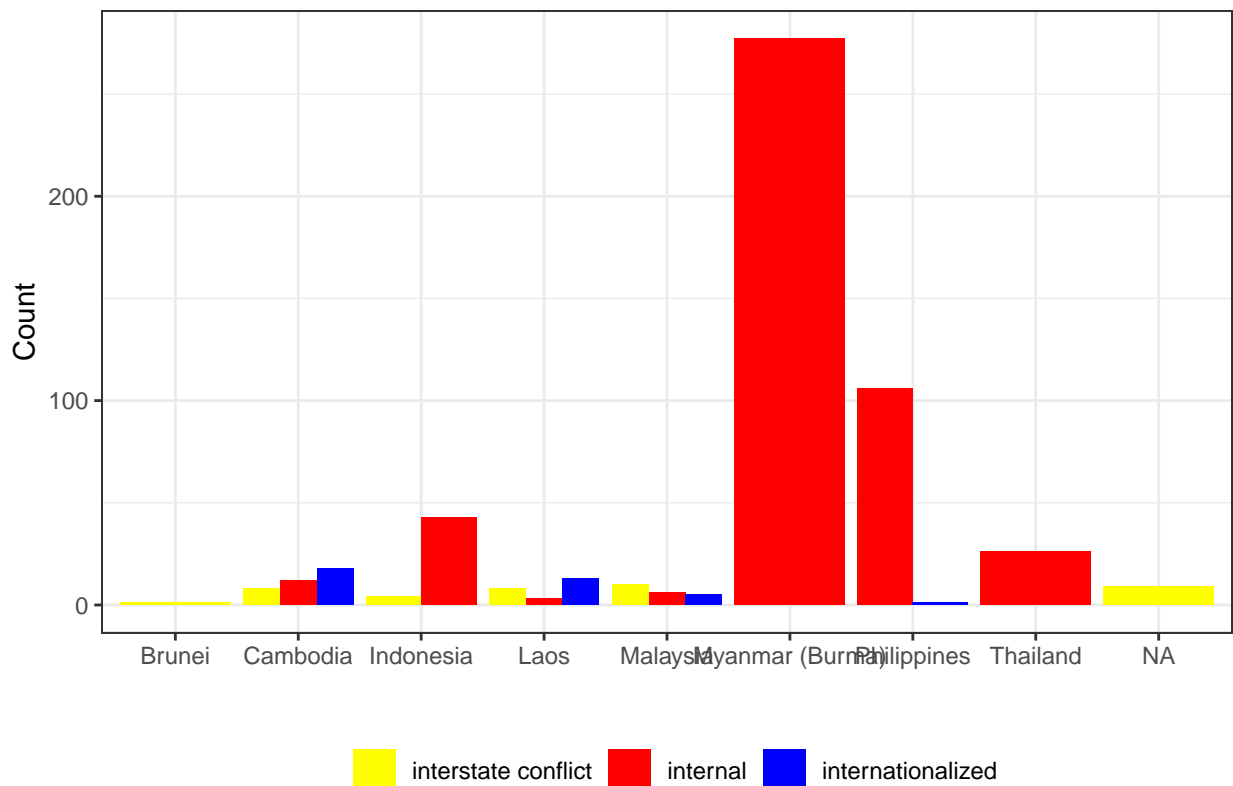
图 8: 东南亚国家抗议与骚乱变化。

- 印度尼西亚、缅甸和柬埔寨以及泰国总体上相对更多的抗议和骚乱，尤其是印度尼西亚的

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。

```
ggplot(ucdp_cnty, aes(fill = type_of_conflict2, y = conflict, x = countyname)) +
  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'))
```



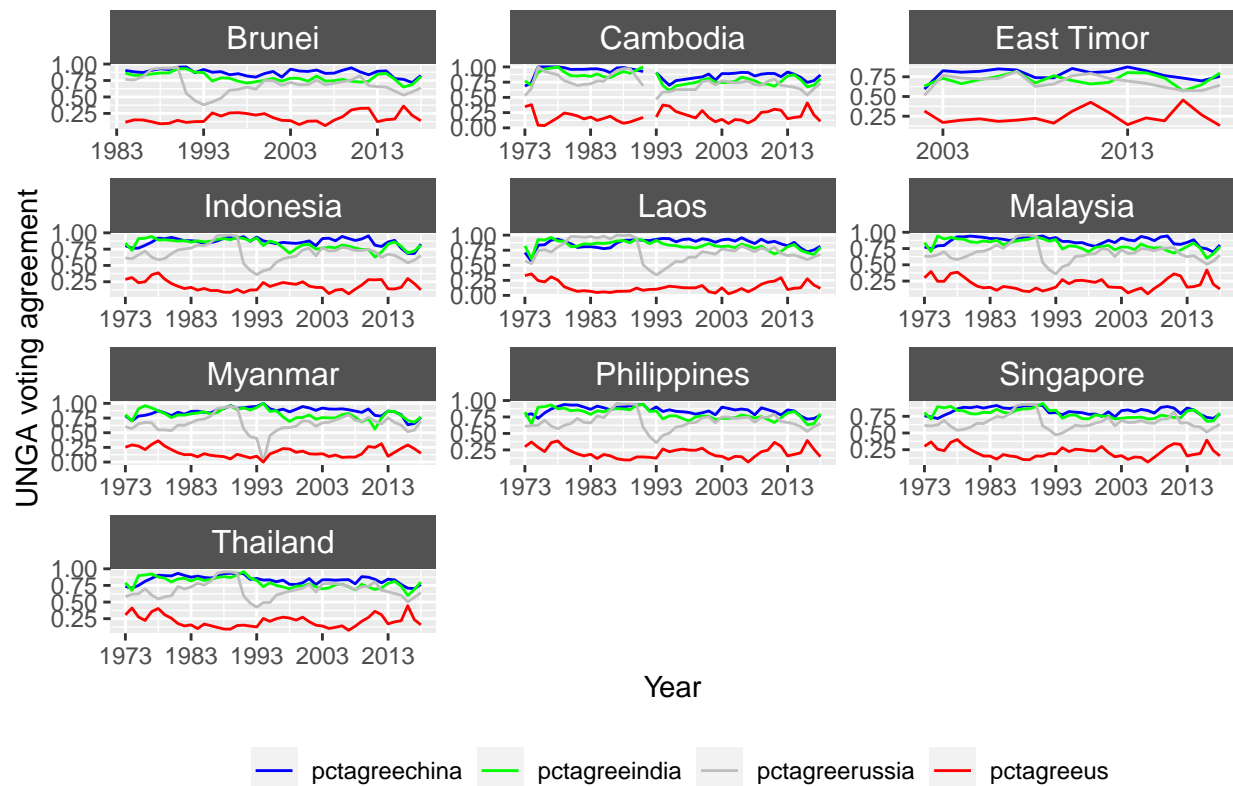
Source: UCDP 2019

图 9: 东南亚各国冲突数量 (1946-2019)。

- 缺失值 (NA): 南越与北越, 1965-1975

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。

```
ggplot(ideal_point_wide, aes(year, agreement, group = type), color= type) +
  geom_line(aes(color=type)) +
  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(~countryname, scales = "free", nrow = 4, ncol = 3) +
  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"))
```



Source: Bailey, Strezhnev, and Voeten (2017, JCR)

图 10: 东南亚各国与主要大国的关系。

- 这一变量取值为 $[0, 1]$, 值越大, 表示越同意该国的立场。
- 东南亚国家在 UN 投票中反映出的总体立场是更同意中国, 其次是印度、俄罗斯 (苏联), 最后是美国