

# 一带一路对中国和其他沿线国家的影响及政策分析

## ——数据科学的视角

范皓年 邓睿哲 李润泽

Peking University

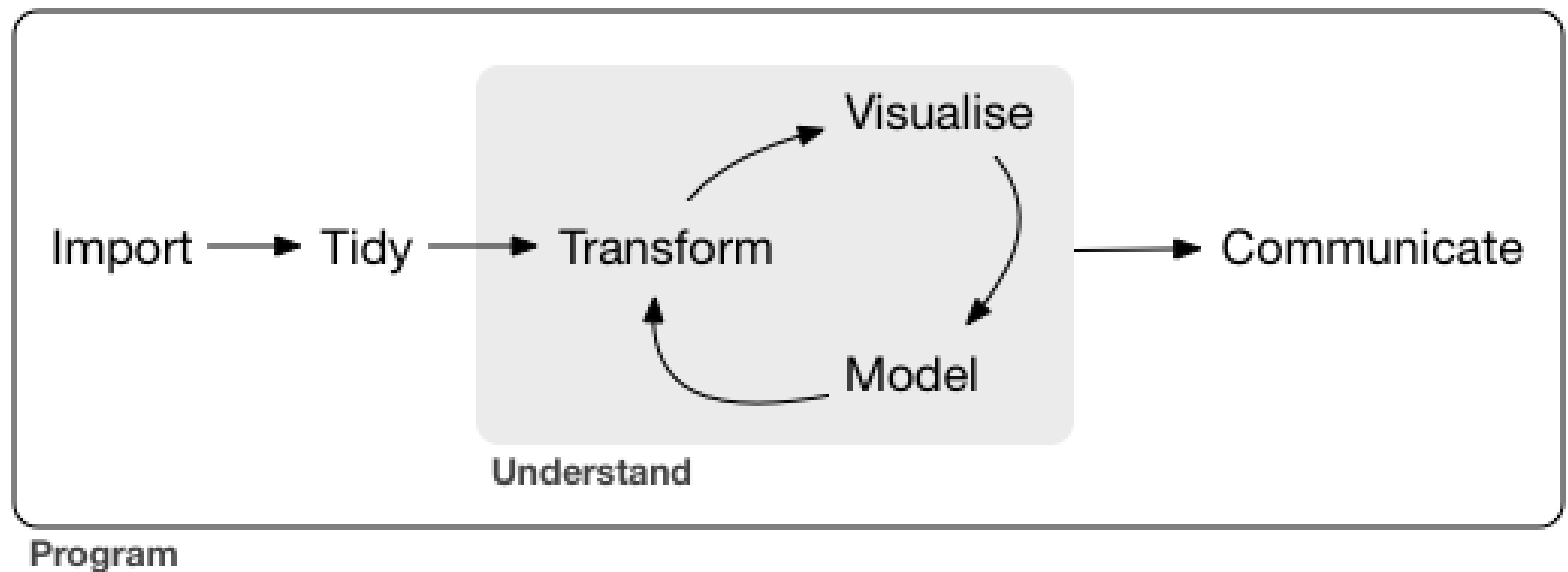
2021-07-06

# 概要

# 主要工作

- The Workflow
- 数据集说明
- 数据清洗
- 数据分析
- 数据可视化

# The Workflow



此图取自 *R for Data Science*，以CC BY-NC-ND 3.0 US发布。

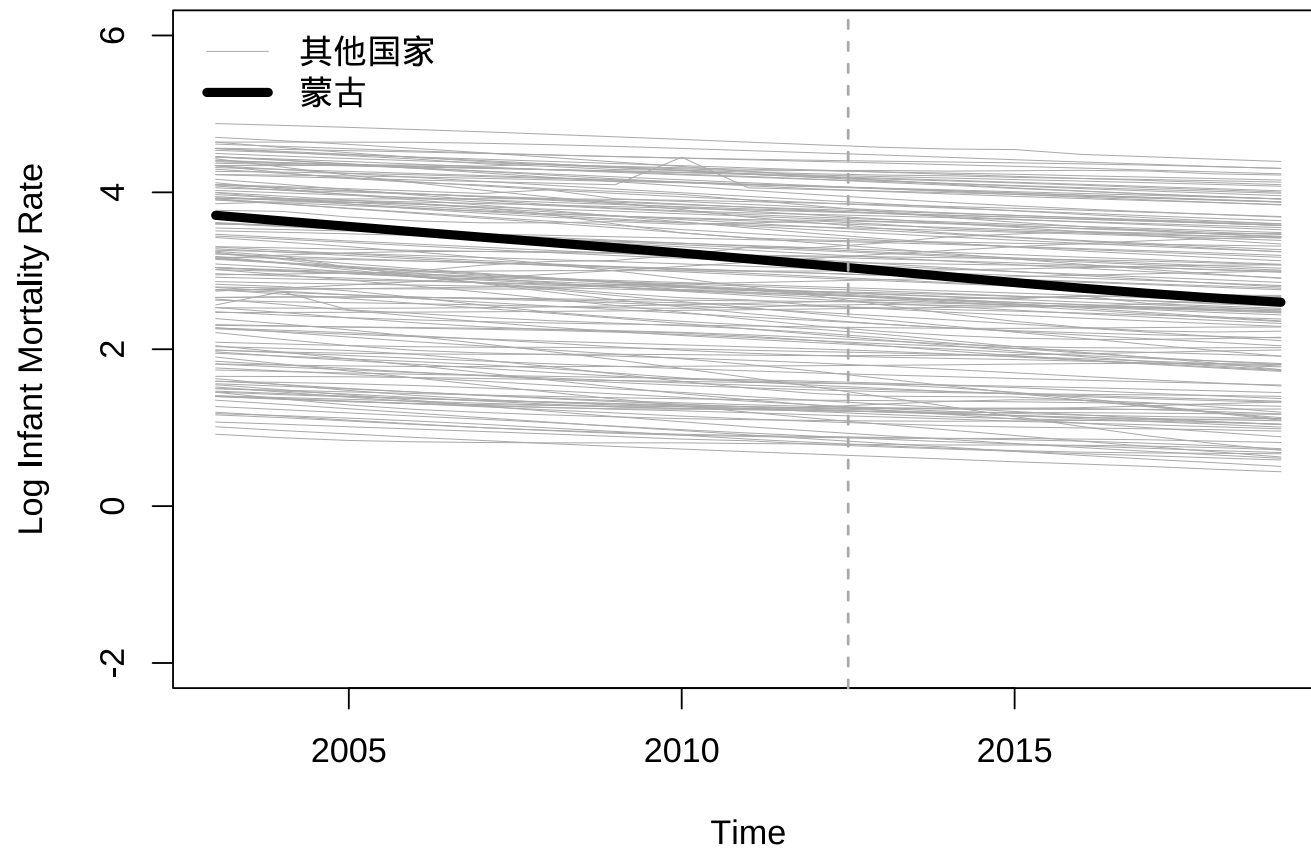
# 数据集说明

国际贸易数据 (`/data/investment/FDI_untidy.csv`) 下载自CEIC数据库. 我们引用CEIC全球数据库中“实际利用的外国资本：按地区分类”和“对外直接投资：国别”两个数据集，逐条下载了中国在利用其他国家资本量（月度数据），和中国对其他国家的直接投资（月度数据）。CEIC数据库覆盖的时间范围为1985年12月至2021年4月，每条数据均以月为统计单位。

其他用到的数据集来自世界健康数据集，包括：

- `under5MortalityRate.csv` 数据集记录了1962-2019年不同国家5岁以下儿童死亡率（每千人的死亡人数）。
- `infantMortalityRate.csv` 数据集记录了1962-2019年不同国家出生婴儿死亡率（每千人的死亡人数）。

# 画个图



# 数据清洗

- 什么样子的数据结构才算是整理好？
- 每行应该代表一个观察(observation)，每列应该代表一个变量(variable).
- 利用tidyverse，结合**正则表达式**，进行数据清洗. 其部分步骤如下页所示.

# 数据清洗

```
simplified_df <- raw_df %>%  
  filter(X1 %>% str_detect("^\\d"))
```

```
fliped_df <- simplified_df %>%  
  pivot_longer(c(-时间), names_to = "observation", values_to = "val")
```

```
str %>%  
  str_replace(pattern = "(.*):(总计|一带一路)", replacement = "1/\\2/\\2")  
  str_replace(pattern = ":::", replacement = ":") %>%  
  str_replace(pattern = "(.*):(.*洲):*(.*)", replacement = "1/\\2/\\3")
```

```
df <- fliped_df %>%  
  mutate(observation = observation %>% stdize()) %>%  
  separate(col = "observation", into = c("type", "地区", "国家"), sep =  
    spread(key = "type", value = "val")
```



# 数据分析

数据建模和分析是传统上受重视的技术.

- 本项目评估了：对测试进行安慰剂检验 (placebo test) 的情况 (`sens.csv`)，测试的P值 (`p.noeff.csv`)，结果的置信区间 (`ci.csv`)
- 并以pdf文件，绘制出了按国家分类的置信区间的情况

# 数据分析

- 缺失数据填补

`mice`: linear regression with bootstrap

- 反事实推断

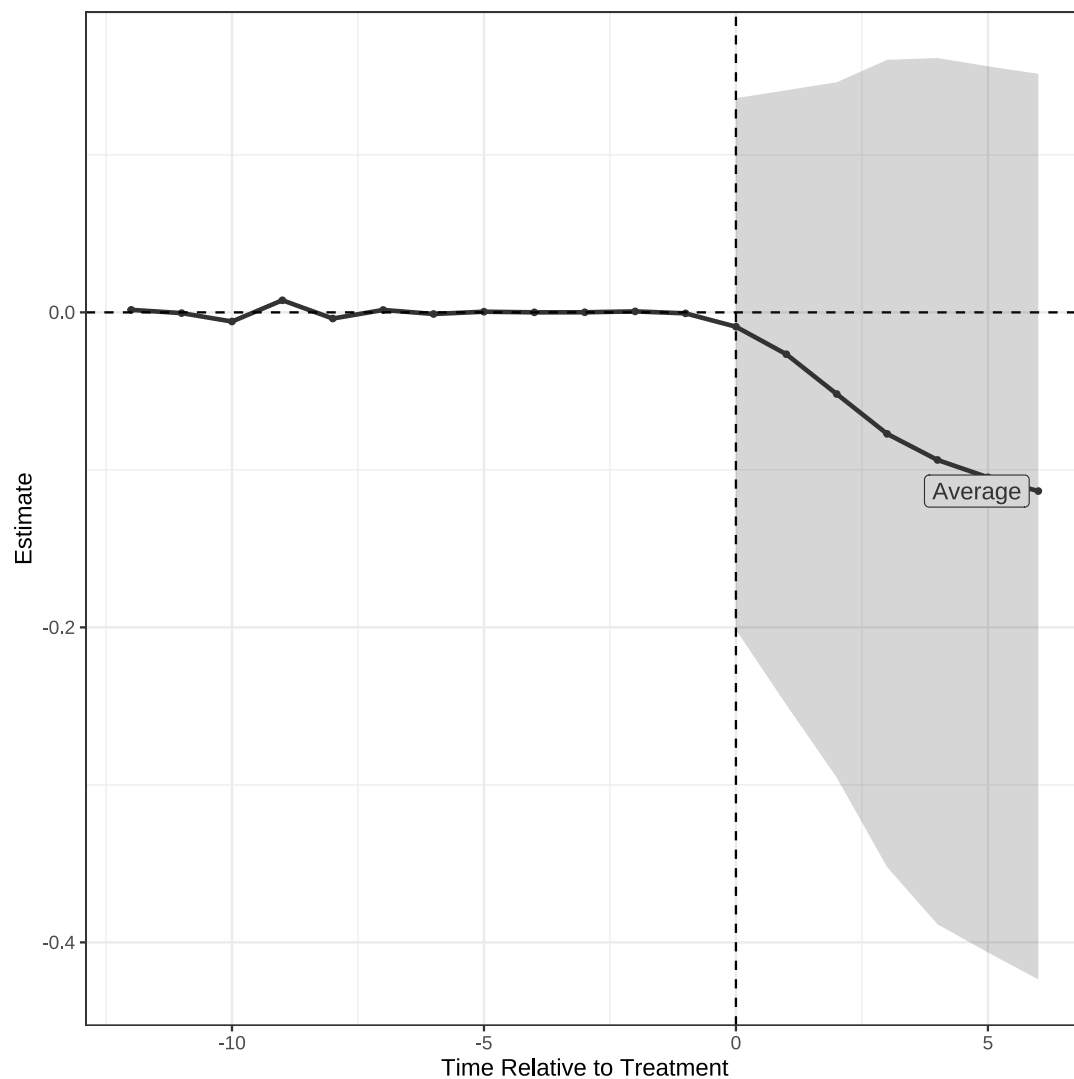
二重差分、合成控制.

$$\hat{P}_t^N = \sum_{j=2}^{J+1} \hat{w}_j Y_{jt}^N$$

- 总体反事实推断

`augsynth`: `multisynth`

## 一带一路对五岁以下儿童死亡率的总效应示意图



# 数据可视化

## 可视化工具

项目将世界经济及其相关的数据，展示在世界地图上，考虑Python语言相对于JavaScript具有更好的数据处理能力，我们使用基于Apache Echarts的Pyecharts：

- 将2003到2019年的中国对外直接投资总额表示在地图上
- 将世界健康数据集中预期寿命和5岁以下死亡率分性别表示在图中

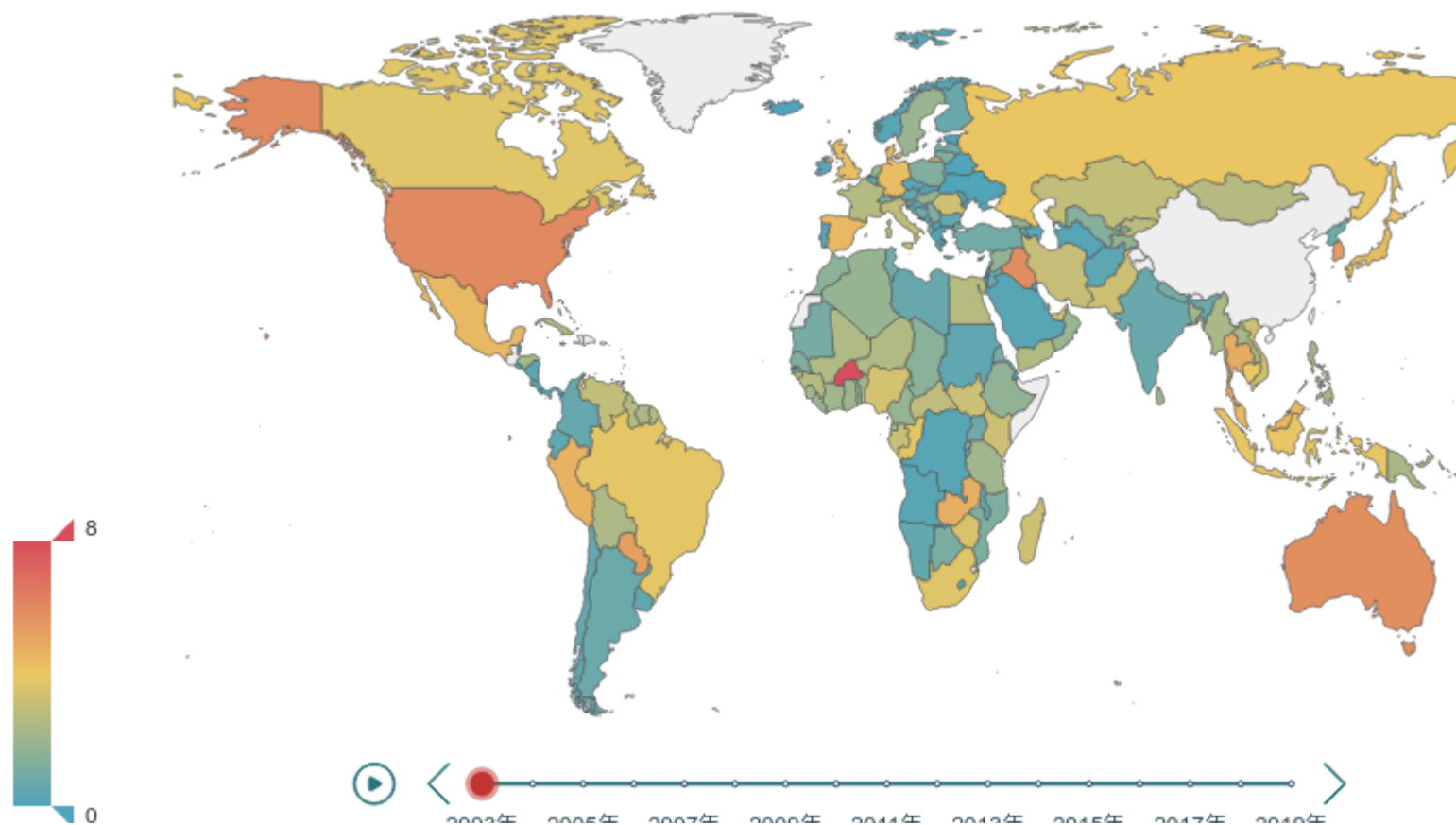
我们从图中可以定性地看出中国外企对于一带一路沿线国家的投入，以及相应国家的经济水平、生活水平的优化.

此外，我们将两种方法生成的置信区间的均值绘图，分别按照时间组织为Timeline排列在Page中. 粗略地我们可以看到，一些特定的年份中，自我国西部延伸至西亚，较高的分层设色部分都分布在一带一路沿线.

# 可交互图形 by ECharts

2003年外商直接投资情况

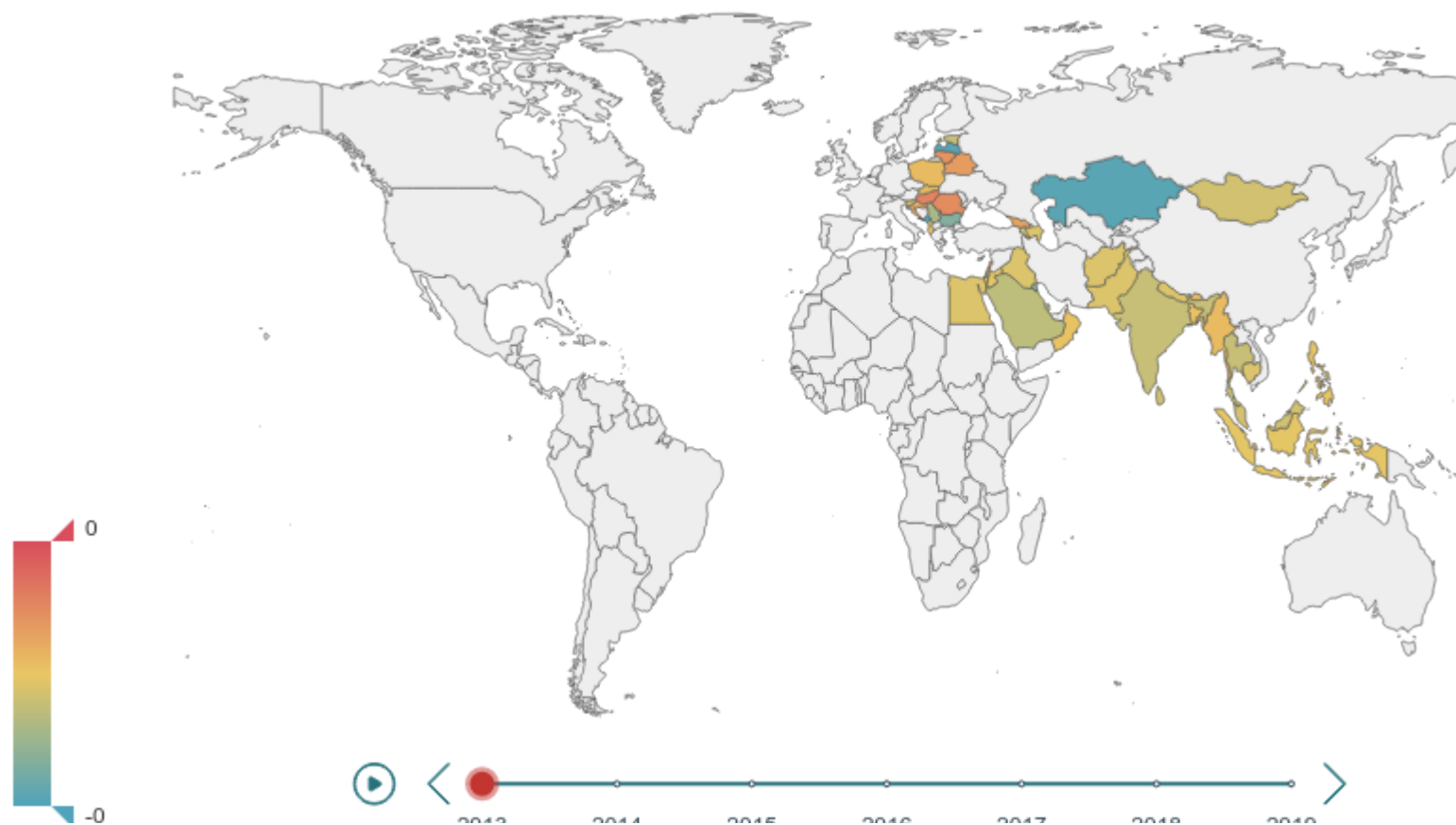
■ 对外直接投资（对数值，原单位：百万美元）



# 可交互图形 by ECharts

2013年外健康状况增长效应 (SC)

置信区间均值



# 总结

# 分析结果

在一带一路的参与国家中，婴儿死亡率相比预期降低（ $P < 0.10$ ）的国家共有4个，如下所示.

```
read_csv("../results/IMR/ci.csv") %>%  
  filter(`max(ci.sc)`<0) %>% group_by(国家) %>% nest() %>% .[[ "国
```

```
## [1] "哈萨克斯坦" "斯洛文尼亚" "拉脱维亚" "黑山"
```

与预期不符，在一带一路的参与国家中，中国对当地的投资相比预期显著降低（ $P < 0.10$ ）的国家仅有1个，如下所示.

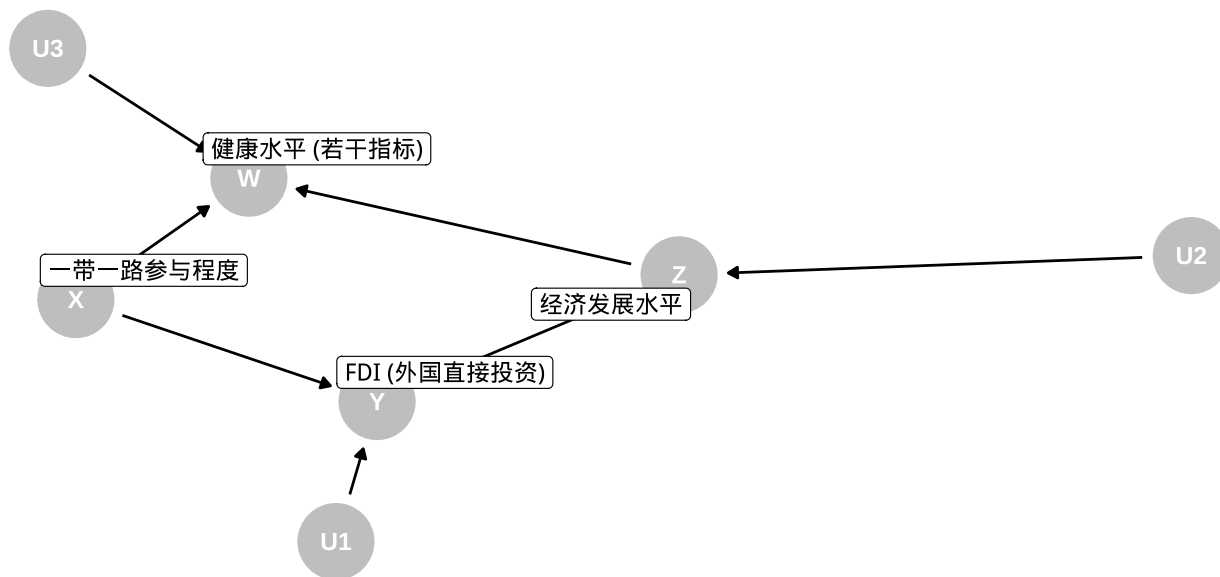
```
read_csv("../results/investment/ci.csv") %>%  
  filter(`min(ci.sc)`>0) %>% group_by(国家) %>% nest() %>% .[[ "国
```

```
## [1] "印度尼西亚"
```

且平均而言，参与一带一路反而额外降低中国对当地的投资的对数约 0.40.



# 分析结果



根据常识，投资(  $Y$  )对经济发展水平(  $Z$  )的平均因果作用是正的，而经济发展水平(  $Z$  )对健康水平(  $W$  )的平均因果作用也是正的。我们建立以上模型。

- 在相同的经济发展水平下，参与一带一路能够相比预期增加沿线国家的国民健康水平，或者狭义来说，**降低婴儿死亡率**。

# 不足和展望

# THANKS

R Markdown syntax,  
Powered by **xaringan** and remark.js