Tourism ~ Year R-squared

1. 加载包，读入数据

```
library(tidyverse)
library(skimr)
raw <- read_csv("outbound_tourism_china.csv")
```

2. 查看数据概况

```
skim(raw)
```

3. 查看各种统计方式包含的国家数

```
raw$SERIES %>% table()

## .
## TCEN TCER  TFN  TFR THSN THSR  VFN  VFR
##   10   23   40   52   13   29   27   28
```

由列联表可看出，TFR 统计方法包含的国家最多，所以选择 TFR 统计方法进行建模

4. 选择 TFR 统计方式，筛选出 TFR 统计方式对应的国家

清洗完数据形式：

```
head(raw1, 4)

## # A tibble: 4 x 23
##   country    `1995` `1996` `1997` `1998` `1999` `2000` `2001` `2002`
`2003`
##   <chr>       <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
 <dbl>
## 1 Angola        NA     NA     NA    215    268    475    653    712
    NA
## 2 Antigua …     NA     NA     NA     NA     NA     NA     NA     NA
    NA
## 3 Armenia        6     35    105     68    158    172    225    305
   345
## 4 Bahamas       NA     NA     NA     NA     NA    503    356    155
   279
## # ... with 13 more variables: `2004` <dbl>, `2005` <dbl>, `2006` <db
l>,
## #   `2007` <dbl>, `2008` <dbl>, `2009` <dbl>, `2010` <dbl>, `2011` <
dbl>,
## #   `2012` <dbl>, `2013` <dbl>, `2014` <dbl>, `2015` <dbl>, `2016` <
dbl>
```

生成国家列表

```r
cnty <- as.list(raw1[,1])$country %>% as.list()
```

将数据清洗为列表函数

```r
tidy_list <- function(country) {
  num <- match(country, cnty)
  dat <- data.frame(year = 1995:2016, tour_num = t(raw1[num, -1]))
  return(dat)
}
```

将数据清洗为包含 52 个国家信息的列表

```r
dat <- map(cnty, tidy_list)
```

创建建立线性模型，并取出 r.squared 函数

```r
mode <- function(x) {
  # 线性回归模型
  mod <-  lm(data = x, formula = tour_num ~ year)
  # 模型摘要
  smy <- summary(mod)
  # 取出摘要中r.squared
  r <- smy[["r.squared"]]
  return(r)
}
```

输出 r.squared，country

```r
r <- lapply(dat, mode) %>% unlist()
country <- cnty %>% unlist()
```

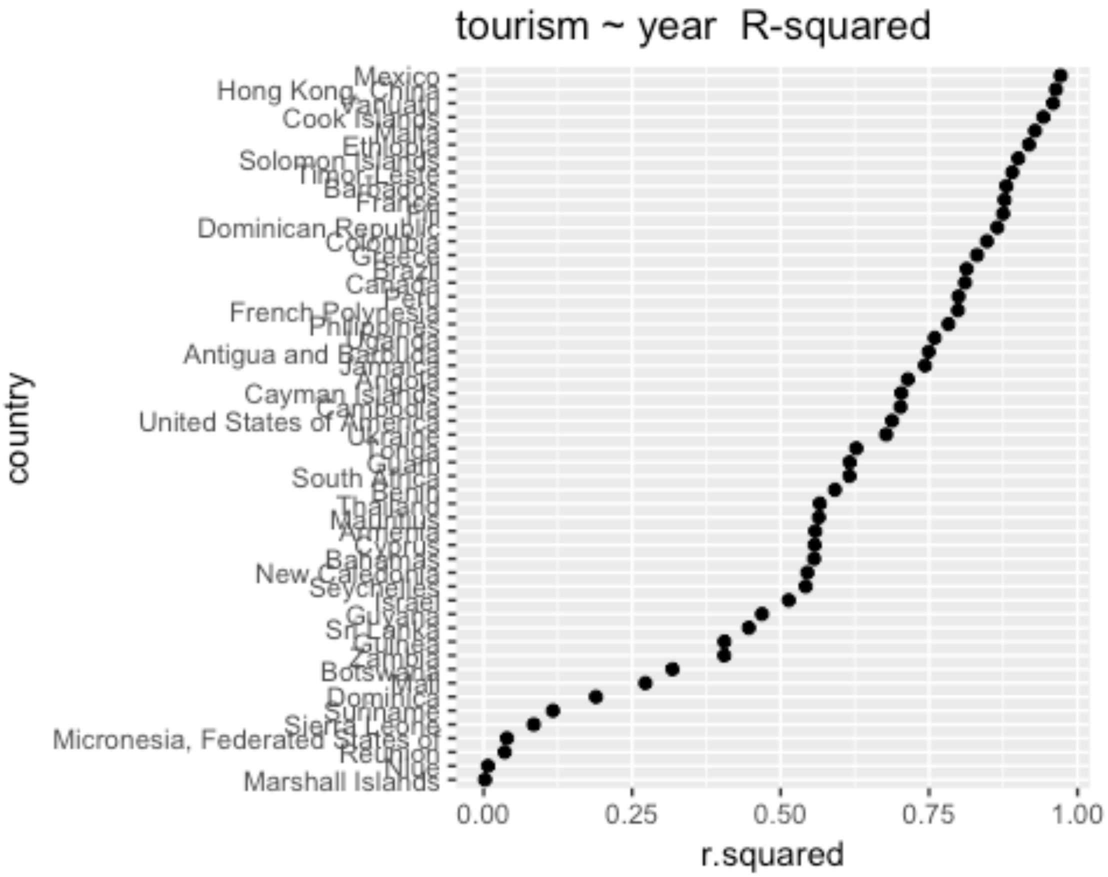
输出包含 country，r.squared 的数据框并按 r.squared 排序

```r
df <- data.frame(country, r.squared = r, stringsAsFactors = F) %>%
  arrange(r.squared)
head(df)

##                            country   r.squared
## 1                 Marshall Islands 0.002498440
## 2                             Niue 0.007384143
## 3                          Reunion 0.035511897
## 4 Micronesia, Federated States of 0.039322964
## 5                     Sierra Leone 0.084339094
## 6                         Suriname 0.116656104
```

输出散点图

```r
ggplot(df, aes(x = r.squared, y = reorder(r.squared, country))) +
  geom_point() +
  scale_y_discrete(breaks = reorder(df$r.squared, df$country),
                         labels = df$country) +
  labs(title = "tourism ~ year  R-squared", x = "r.squared", y = "count
ry")
```


tourism ~ year R-squared
country
Mexico
Hong Kong, China
Vanuatu
Cook Islands
Malta
Ethiopia
Solomon Islands
Timor-Leste
Barbados
France
Fiji
Dominican Republic
Colombia
Greece
Brazil
Canada
Peru
French Polynesia
Philippines
Uganda
Antigua and Barbuda
Jamaica
Angola
Cayman Islands
Cambodia
United States of America
Ukraine
Tonga
Guam
South Africa
Benin
Thailand
Mauritius
Armenia
Cyprus
Bahamas
New Caledonia
Seychelles
Israel
Guyana
Sri Lanka
Guinea
Zambia
Botswana
Mali
Dominica
Suriname
Sierra Leone
Micronesia, Federated States of
Reunion
Niue
Marshall Islands
r.squared