**Assignment 02**

# 1. Significant earthquakes since 2150 B.C.

1.1 读取signif.txt至Sig\_Eqs

Data\_E <- read.delim("signif.txt",head=TRUE)

Sig\_Eqs <- as\_tibble(Data\_E)

Sig\_Eqs

1.2 选取YEAR, COUNTRY, DEATHS，利用group\_by函数，集中对COUNTRY进行排列整理，利用summarize和sum函数计算总的死亡人数，最后用arrange函数进行排序**。**

**select(Sig\_Eqs, YEAR, COUNTRY, DEATHS) %>%**

**group\_by(COUNTRY) %>%**

**summarise(countrytotal=sum(DEATHS, na.rm = T)) %>%**

**arrange(desc(countrytotal))**

****

图1

1.3 选取YEAR, EQ\_PRIMARY，利用filter函数筛选出EQ\_PRIMARY >= 6.0的数据，再利用group\_by函数，集中对YEAR进行排列整理，count函数进行记数求取每年地震强度大于6的次数，最后plot——图2。

select(Sig\_Eqs, YEAR, EQ\_PRIMARY) %>%

filter(EQ\_PRIMARY >= 6.0) %>%

group\_by(YEAR) %>%

count() %>%

ggplot(aes(x=YEAR, y=n)) +

geom\_point()

或者，在利用filter函数筛选出EQ\_PRIMARY >= 6.0的数据之后，直接plot，根据每年点的密集程度判断次数——图3。

select(Sig\_Eqs, YEAR, EQ\_PRIMARY) %>%

filter(EQ\_PRIMARY >= 6.0) %>%

ggplot(aes(x=YEAR, y=EQ\_PRIMARY)) +

geom\_point()

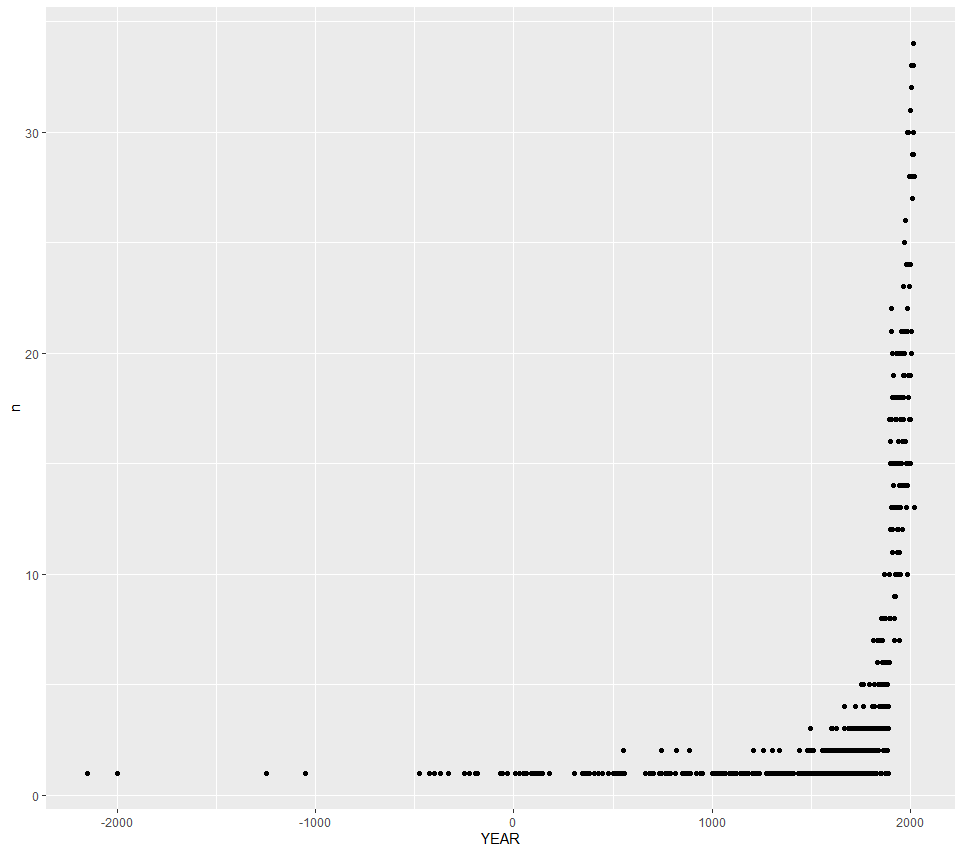


图2

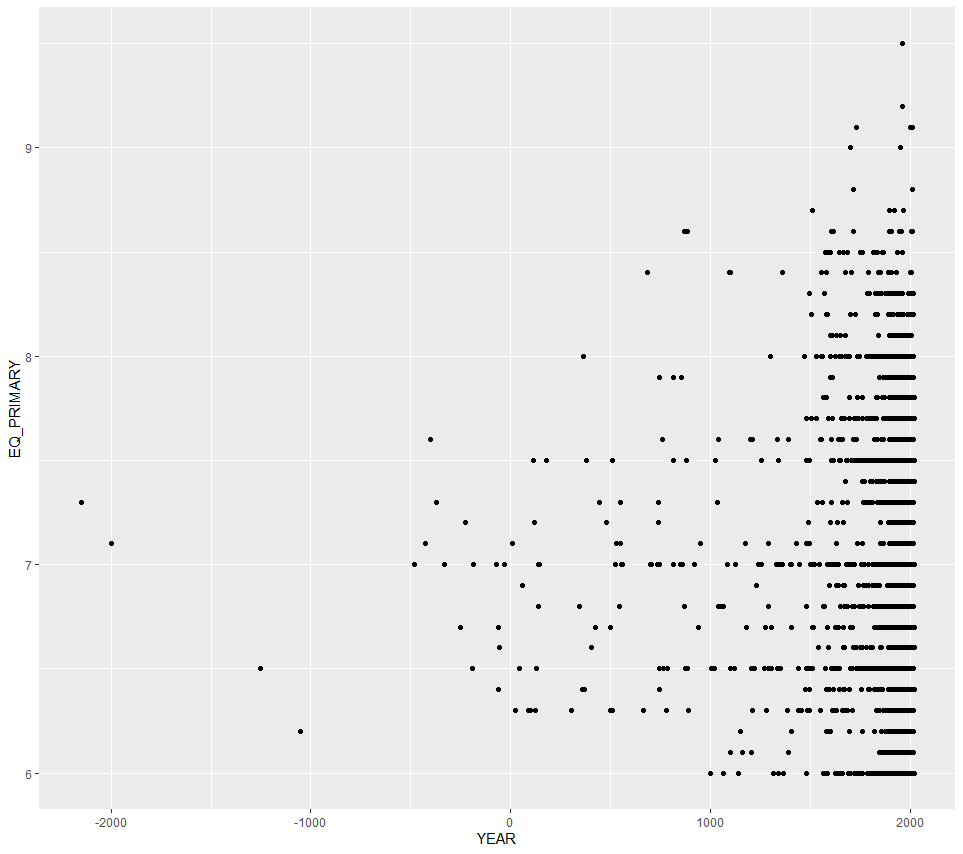


图3

结果与讨论：随着时间向近代推进，强度大于6的地震发生的次数变多，且地震的最大强度、振幅值也变大，从两个角度分析其原因，首先，第一点，可能是由于板块运动在近代更加活跃造成的；其次，第二点，就是人类活动的影响，例如大型水库工程等建设，改变了原有板块的应力状态，造成地震频发，以及大强度地震的发生。

1.4 先计算每个国家发生地震的总数，计入country\_total。

country\_total <- Sig\_Eqs %>%

group\_by(COUNTRY) %>%

count()

然后整理筛选出每个国家历史发生的最大地震，计入country\_max。

country\_max <- Sig\_Eqs %>%

select(YEAR, MONTH, DAY, COUNTRY,EQ\_PRIMARY) %>%

mutate(DATE=paste(YEAR,MONTH,DAY,sep = "\_")) %>%

select(DATE, COUNTRY, EQ\_PRIMARY) %>%

group\_by(COUNTRY) %>%

summarize(magmax=max(EQ\_PRIMARY,na.rm = T))

接着对日期进行整理合并，筛选出日期对应国家及地震强度计入country\_date。

country\_date <- Sig\_Eqs %>%

select(YEAR, MONTH, DAY, COUNTRY,EQ\_PRIMARY) %>%

mutate(DATE=paste(YEAR,MONTH,DAY,sep = "\_")) %>%

select(DATE, COUNTRY, EQ\_PRIMARY)最后建立CountEq\_LargestEq函数。

CountEq\_LargestEq <- function(states){

total\_C <- country\_total %>%

filter(COUNTRY==states)

max\_magnitude <- country\_max %>%

filter(COUNTRY==states)

max\_magnitude2 <- max\_magnitude$magmax

max\_date <- country\_date %>%

filter(EQ\_PRIMARY==max\_magnitude2 & COUNTRY==states)

print("The total number of earthquakes since 2150 B.C.is ")

print(total\_C$n)

print("the date of the largest earthquake ever happened in this country is ")

print(max\_date$DATE)

}

根据国家筛选出该国家发生地震的次数以及最大地震发生的日期。

> CountEq\_LargestEq("CHINA")

[1] "The total number of earthquakes since 2150 B.C.is "

[1] 606

[1] "the date of the largest earthquake ever happened in this country is "

[1] "1668\_7\_25"

但是，这部分稍微有一点问题，不能显示最大强度地震多次发生的时间，且不能实现后面要做的排序，因此，在此基础之上，对这一小问的语句进行修改，以实现后面要做的排序，见PS2\_1-2.R文件中1.4的语句，如下，建立CountEq\_LargestEq函数，刷选出某国家地震总次数以及最大强度地震对应的时间，并计入max\_data，与上一个语句的区别是，这里将结果以data.frame的形式显示，且能够显示该国家最大强度地震多次发生的时间。

CountEq\_LargestEq <- function(states){

# total\_C <- country\_total %>%

# filter(COUNTRY==states)

max\_magnitude <- country\_max %>%

filter(COUNTRY==states)

max\_magnitude2 <- max\_magnitude$magmax

max\_date <- country\_date %>%

group\_by(COUNTRY) %>%

mutate(count= n()) %>%

filter(EQ\_PRIMARY==max\_magnitude2 & COUNTRY==states)

#print("The total number of earthquakes since 2150 B.C.is ")

# print(total\_C$n)

# print("the date of the largest earthquake ever happened in this country is ")

# print(max\_date$DATE)

max\_date

}

CountEq\_LargestEq("CHINA")

接着，以所有的国家名字做for循环，利用mutate筛选出将所有国家对应的地震总次数以及最大强度地震对应的时间，计入Result，最后对Result做arrange排序。

country <- country\_total$COUNTRY

Result <- c()

for (i in country) {

max <- CountEq\_LargestEq(i)

Result <- rbind(Result, max)

}

Order <- Result %>%

arrange(desc(count))

Order

# A tibble: 162 x 4

# Groups: COUNTRY [144]

DATE COUNTRY EQ\_PRIMARY count

<chr> <chr> <dbl> <int>

1 1668\_7\_25 CHINA 8.5 606

2 2011\_3\_11 JAPAN 9.1 408

3 2004\_12\_26 INDONESIA 9.1 394

4 856\_12\_22 IRAN 7.9 381

5 1912\_8\_9 TURKEY 7.8 329

6 1916\_1\_24 TURKEY 7.8 329

7 1915\_1\_13 ITALY 7.5 326

8 1964\_3\_28 USA 9.2 268

9 365\_7\_21 GREECE 8 264

10 1303\_8\_8 GREECE 8 264

# ... with 152 more rows

# 2. Wind speed in Shenzhen during the past 10 years

读取2281305.csv至SZ\_wind。建立montly\_mean函数，记录每一年每一个月的平均风速，最后plot——图4。

Met\_Data <- read.csv(file = "2281305.csv", header = T)

names(Met\_Data)

SZ\_wind <- as\_tibble(Met\_Data)

SZ\_wind

montly\_mean <- c()

for (i in 2010:2020) {

Thisyear\_mean <- SZ\_wind %>%

select(DATE, WND) %>%

mutate(wind\_speed=substr(WND, 9, 12)) %>%

mutate(wind\_speed2=as.numeric(wind\_speed)) %>%

mutate(month=substr(DATE,1,7)) %>%

#mutate(month2=as.numeric(month)) %>%

mutate(year=substr(DATE,1,4)) %>%

mutate(year2=as.numeric(year)) %>%

filter(year2==i) %>%

group\_by(month) %>%

summarise(mean=mean(wind\_speed2, na.rm = T))

montly\_mean <- rbind(montly\_mean, Thisyear\_mean) #help from TA Ming YANG

}

montly\_mean %>%

ggplot(aes(x=month, y=mean))+

geom\_point()

#需要改进横坐标的显示

#TA Ming YANG提供了一些帮助

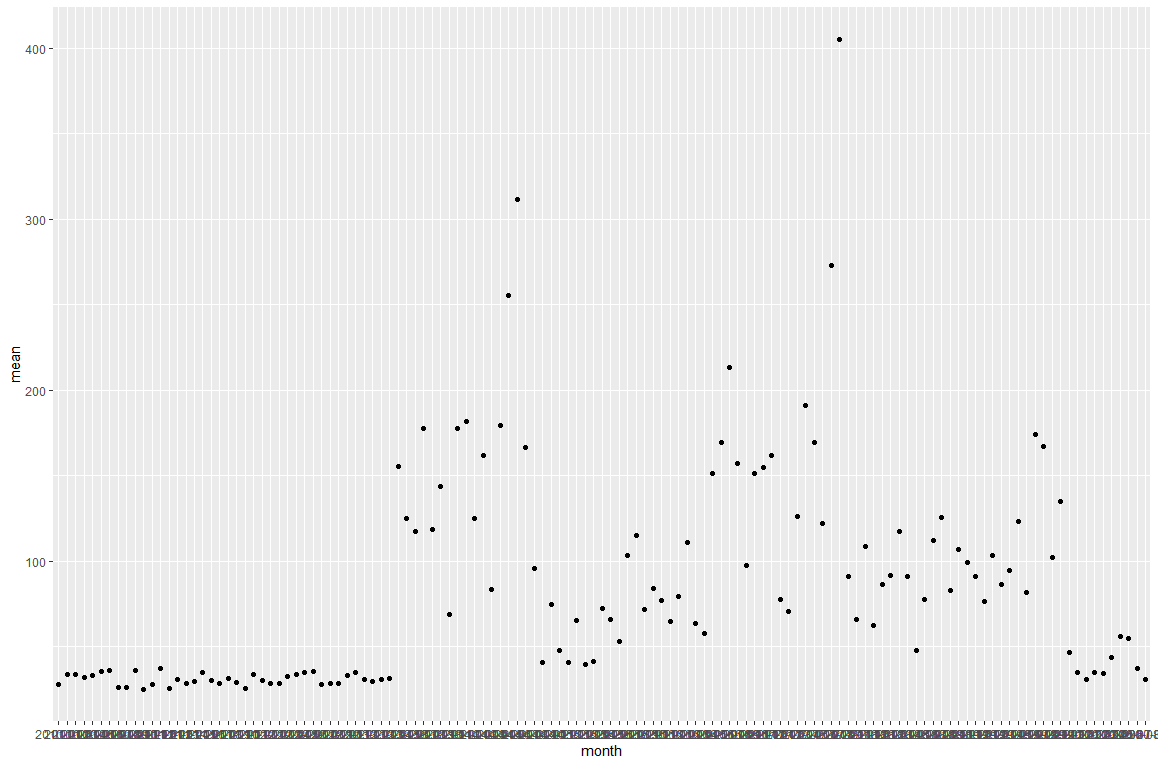


图4

结果与讨论：前几年风速还比较平稳，而近些年，风速的变化较大。推测主要是与人类活动有关，随着经济高速发展，地表建筑等发生了较大的变化，其次是耕地与林地面积的变化也是造成风速变化的另一个原因。

# 3. Revisit a data set

读取VQC00670480.csv至Pre\_Data

Precipitation\_Data <- read.csv(file = "VQC00670480.csv", header = T)

names(Precipitation\_Data)

Pre\_Data <- as\_tibble(Precipitation\_Data)

1. plot降雨与时间图，2016年之后——图5。利用filter和mutate对数据进行筛选和处理。

select(Pre\_Data, DATE, DlySum, DlySumQF) %>%

filter(DATE > "2016-12-31") %>%

mutate(DlySum\_new=ifelse(DlySumQF=="M", NA, DlySum)) %>%

ggplot(aes(x=DATE, y=DlySum\_new)) +

geom\_point()

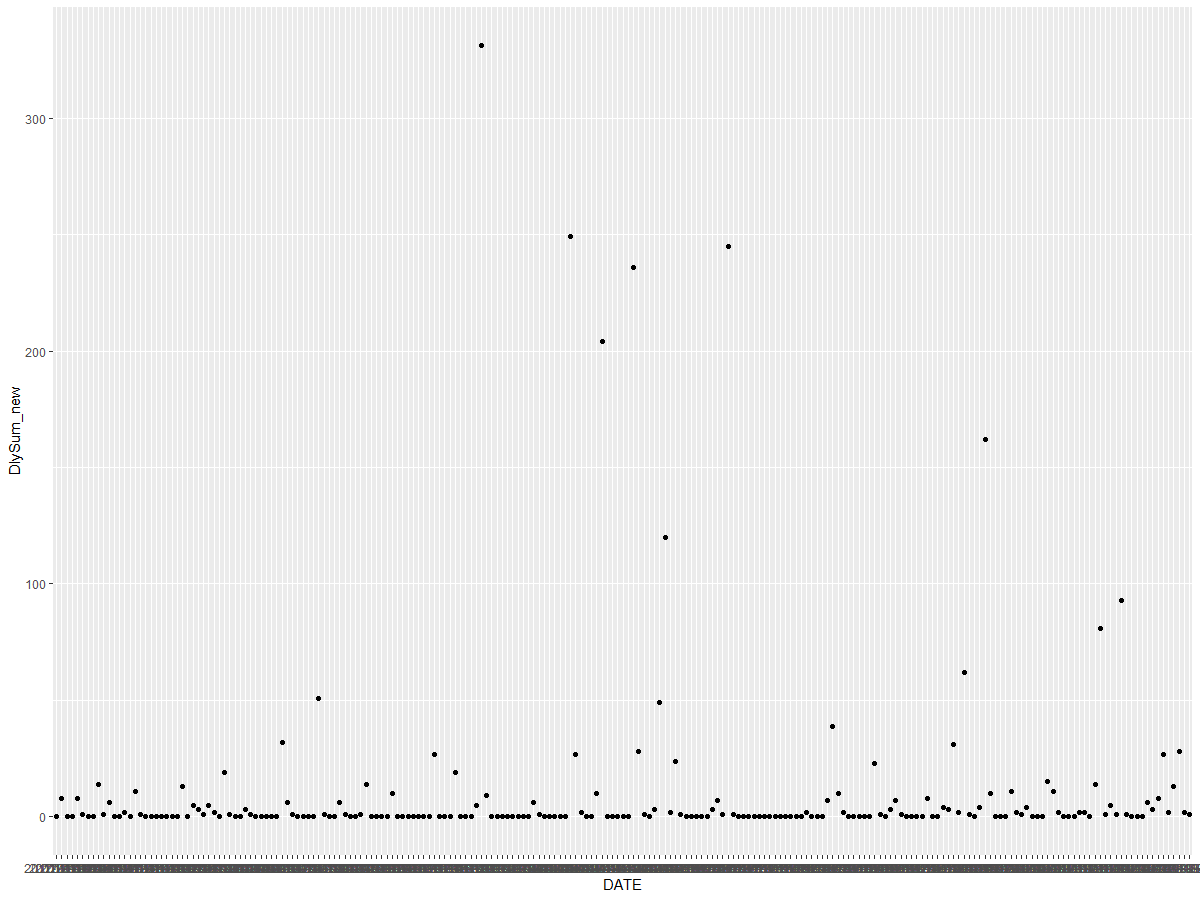


图5

2. 计算每年降雨的平均值、最大最小值等计算——图6。

select(Pre\_Data, DATE, DlySum, DlySumQF) %>%

mutate(Year=substr(DATE,1,4)) %>%

filter(DlySumQF!="M") %>%

group\_by(Year) %>%

summarize(Annual\_mean=mean(DlySum), Annual\_sd=sd(DlySum),

Annual\_min=min(DlySum),Annual\_max=max(DlySum),

Annual\_se=sd(DlySum)/sqrt(n()))

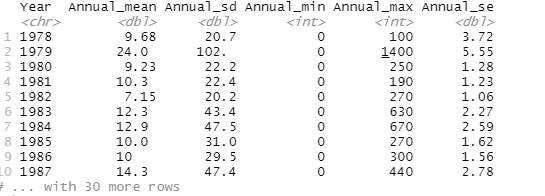


图6