

1 Significant earthquakes since 2150 B.C.

- (1) 从 CSDN 上阅读“R 语言-文本文件读写 txt/csv/xlsx”： csv 并不是一种单一的、定义明确的格式，常见的空白分隔符有空格（sep=“ ”）、制表符和换行符，根据题意使用其中的制表符为 sep=“\t”，然后读取“signif.txt”文件：

```
library(tydyr)
library(dplyr)
library(ggplot2)
EQ_D<-read.csv(file="signif.txt",sep = "\t",header = T)
Sig_Eqs=as_tibble(EQ_D)
```

- (2) 先选出与题意有关的数据：国家、死亡总数，通过 country 分组，用 summarize() 计算每个国家的死亡总数，然后用 arrange(desc()) 将死亡总数从大到小排列，run 以后会显示数据最大的前十行。

```
#1.2
Sig_Eqs %>%
  select(COUNTRY,TOTAL_DEATHS) %>%
  group_by(COUNTRY) %>%
  summarize(EachC_total_death=sum(TOTAL_DEATHS,na.rm = T)) %>%
  arrange(desc(EachC_total_death))
```

- (3) 计算每年全球范围内震级大于 6.0（使用 EQ_PRIMARY 作为震级）的地震总数，先用 select() 年份和 EQ_PRIMARY，再用 filter() 选出震级大于 6.0 的，按照年分组然后画图。

```
#1.3
Sig_Eqs %>%
  select(YEAR,EQ_PRIMARY) %>%
  filter(EQ_PRIMARY>6.0) %>%
  group_by(YEAR) %>%
  summarize(total_per=sum(EQ_PRIMARY,na.rm = T)) %>%
  ggplot((aes(x=YEAR,y=total_per)))+
  geom_line()
```

#随着年份增长，呈现显著增加的趋势，我认为主要原因是因为科技越来越发达，记录到的地震的次数和情况也越来越全面和详细，以前主要靠人为记录较多，也不排除人类对自然的改造对地质构造产生了影响，使得局部地区构造断裂带较以往相对活跃，但我认为这并不是使地震记录次数显著增长的主要原因

- (4) 题目涉及到国家、地震总数和日期，根据题意，首先建立一个关于国家的函数 CountEq_LargestEq()，先按照自己的思路写了返回该国有史以来最大地震的日期，思路还不是很通畅。

```
#1.4
#返回该国有史以来最大地震的日期
CountEq_LargestEq<-function(country){
  Largest_EQ<-Sig_Eqs %>%
    filter(COUNTRY==country) %>%
    mutate(date=paste(YEAR,MONTH,DAY,sep = '-')) %>%
    select(date,EQ_PRIMARY) %>%
    summarise(Largest_EQ=n(),date_max=
      date[which(EQ_PRIMARY==max(EQ_PRIMARY))]) %>%
  List(Largest_EQ,date_max)
  return(date_max)
}
```

2 Wind speed in Shenzhen during the past 10 years

首先 `library()`, 然后读取数据, 转换数据格式, 选出与题目意思有关的数据, 即 `DATE`、`WND`, 然后根据用户手册 8-9 页按照顺序选出 `WND` 中存在且质量好的数据, 根据用户手册, 风速为 `WND` 数据的第 9-12 位数, 用 `substr()` 选出 9-12 位作为 `speed rate`, 并用 `mutate()` 新增这列数据, 而日期格式为 2020-10-23, 我尝试用此格式画图画不出来:

```
geom_line: na.rm = FALSE, orientation = NA
stat_identity: na.rm = FALSE
position_identity
```

所以为了按月分类画图, 将 `date` 格式的年月数据用 `as.numeric()` 转换为数字格式, 提取出来, 将年和月以第 1 题第 (1) 问提到空白分隔符用 `paste()` 连接起来的并新增一列, 再选出新增的两列, 按照月分组以后计算每月的平均风速, 再以日期为 `x` 轴, 风速为 `y` 轴, 用 `ggplot()` 画图。

```
library(tidyr)
library(dplyr)
library(ggplot2)
wind=read.csv(file="2281305.csv",header = T)
wind.tbl<-as_tibble(wind)
wind.tbl %>%
  select (DATE,WND) %>%
  filter(substr(WND,1,3)!="999") %>%
  filter(substr(WND,9,12)!="9999") %>%
  filter(substr(WND,5,5)=="1") %>%
  filter(substr(WND,7,7)=="N") %>%
  filter(substr(WND,14,14)=="1") %>%
  mutate(speed_rate=as.numeric(substr(WND,9,12))) %>%
  mutate(date_month=as.numeric(paste(substr(DATE,1,4),substr(DATE,6,7),sep = " "))) %>%
  select(date_month,speed_rate) %>%
  group_by(date_month) %>%
  summarize(monthly_mean=mean(speed_rate,na.rm=T)) %>%
  ggplot(aes(x=date_month,y=monthly_mean))+
  geom_line()
```

3 Revisit a data set

首先 `library()`, `read.csv()` 读取数据, `as.tibble()` 转换数据格式, 选出盐度大于 0 的数据, 将日期 (`chr`) 格式转换为 `date` 格式, 盐度(`chr`)格式转换为 `double` 格式, 新增列, 选择新增两列数据, 画图。

```
library(tidyr)
library(dplyr)
library(ggplot2)
sz <- read.csv(file="water.csv",header = T,encoding = "UTF-8")
sz_tbl<-as_tibble(sz)
sz_tbl %>%
  filter(Salinity..psu.>0) %>%
  mutate(date=as.Date(Dates)) %>%
  mutate(S=as.double(Salinity..psu.)) %>%
  select(date,S) %>%
  ggplot(aes(x=date,y=S))+
  geom_line()
```