

Problem1

这一题中需要用到这三个包，所以先导入：

```
library(tidyr)
```

```
library(dplyr)
```

```
library(ggplot2)
```

1.1 signif.txt 为带有制表符格式的文件，读取文件并转换为 tibble

```
Sig_data<-read.csv(file
"C:/Users/Administrator/Documents/signif.txt",header=TRUE,sep="\t")
```

```
Sig_Eqs <- as_tibble(Sig_data)
```

1.2 计算每个国家因地震造成的死亡总数，并打印出前十个国家。先 select 出 year,country,death,然后利用 group-by 加上 summarize 计算出每个国家的死亡总数，再按照降序排列打印出前十个国家的数据。

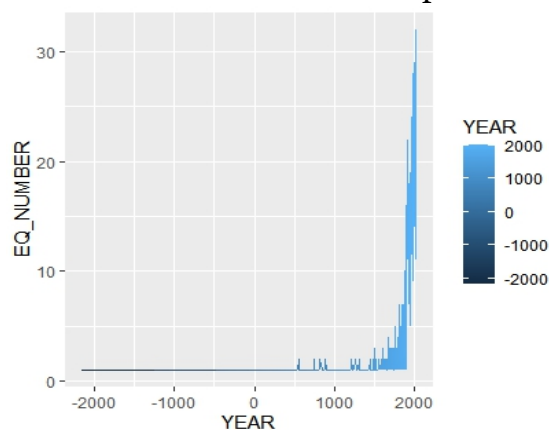
```

  COUNTRY
  <chr>
1 BARBADOS
2 GUINEA
3 LIBYA
4 UGANDA
5 IRELAND
6 WALLIS AND FUTUNA (FRENCH TERRITORY)
7 BURUNDI
8 BELGIUM
9 CZECH REPUBLIC
10 DJIBOUTI
# ... with 145 more rows

```

	a
1 BARBADOS	3000
2 GUINEA	443
3 LIBYA	300
4 UGANDA	152
5 IRELAND	100
6 WALLIS AND FUTUNA (FRENCH TERRITORY)	5
7 BURUNDI	3
8 BELGIUM	2
9 CZECH REPUBLIC	2
10 DJIBOUTI	2

1.3 首先利用 filter 筛出震级大于 6.0 的数据，利用 group-by 加上 summarize 统计出各年份的地震发生的次数，plot 出图。



1.4 (本题得到了李熹成同学的关于用矩阵输出结果的指导)

首先编写函数 CountEq_LargestEq, filter 选出制定某国家以及震级不为 NA 的行数, 利用 mutate 和 paste 将日期列加入进去, 利用 summarize 统计行数及该国家地震发生次数和最大地震对应的日期。将这两列数据导入新的 data moment-data 中。

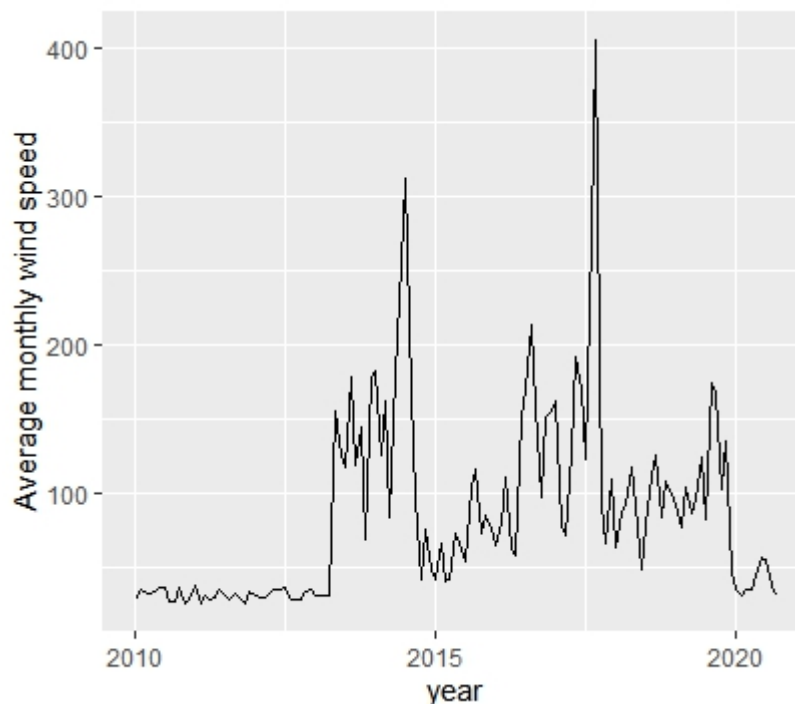
下面利用矩阵将 moment-data 中的结果输出即可。输出时按照地震总次数的降序排列。结果如下 :

	[,1]	[,2]	[,3]
[1,]	"CHINA"	"575"	"1668-7-25"
[2,]	"JAPAN"	"343"	"2011-3-11"
[3,]	"INDONESIA"	"314"	"2004-12-26"
[4,]	"IRAN"	"249"	"856-12-22"
[5,]	"USA"	"215"	"1964-3-28"
[6,]	"TURKEY"	"206"	"1912-8-9"
[7,]	"GREECE"	"152"	"365-7-21"
[8,]	"PERU"	"146"	"1716-2-6"
[9,]	"CHILE"	"145"	"1960-5-22"
[10,]	"RUSSIA"	"139"	"1952-11-4"
[11,]	"PHILIPPINES"	"132"	"1897-9-21"
[12,]	"MEXICO"	"119"	"1899-1-24"
[13,]	"ITALY"	"96"	"1915-1-13"
[14,]	"TAIWAN"	"93"	"1920-6-5"
[15,]	"PAPUA NEW GUINEA"	"89"	"1919-5-6"
[16,]	"INDIA"	"81"	"1950-8-15"
[17,]	"NEW ZEALAND"	"62"	"1826-NA-NA"
[18,]	"SOLOMON ISLANDS"	"60"	"1977-4-21"
[19,]	"COLOMBIA"	"55"	"1826-6-18"
[20,]	"AFGHANISTAN"	"53"	"1909-7-7"
[21,]	"ECUADOR"	"53"	"1906-1-31"
[22,]	"VANUATU"	"48"	"1913-10-14"
[23,]	"PAKISTAN"	"43"	"1945-11-27"
[24,]	"ALGERIA"	"38"	"1980-10-10"
[25,]	"ALBANIA"	"34"	"1893-6-14"
[26,]	"VENEZUELA"	"30"	"1900-10-29"
[27,]	"GUATEMALA"	"28"	"1942-8-6"
[28,]	"NICARAGUA"	"27"	"1898-4-29"
[29,]	"COSTA RICA"	"24"	"1950-10-5"
[30,]	"TAJIKISTAN"	"24"	"1907-10-21"
[31,]	"MYANMAR (BURMA)"	"24"	"1912-5-23"
[32,]	"USA TERRITORY"	"22"	"1902-9-22"
[33,]	"EL SALVADOR"	"21"	"1915-9-7"
[34,]	"AUSTRALIA"	"21"	"1989-5-23"
[35,]	"NEW CALEDONIA"	"20"	"1875-3-28"
[36,]	"PANAMA"	"20"	"1882-9-7"
[37,]	"SOUTH KOREA"	"19"	"1643-7-25"
[38,]	"TONGA"	"18"	"1919-4-30"
[39,]	"KERMADEC ISLANDS (NEW ZEALAND)"	"17"	"1986-10-20"
[40,]	"FIJI"	"17"	"1919-1-1"
[41,]	"ARGENTINA"	"16"	"1944-1-15"

Problem2

我们发现 Data 中 WND 这一列，1-3 为 wind_angle，5 为 wind_dqc (direction quality code)，7 为 wind_tcode (type code)，9-12 为 wind_speed，14 为 wind_speedqc (quality code)。利用 mutate 以及 substr 将各项指标分隔好成列。

我们阅读 user guide 第 8,9 页，开始过滤数据，例如 wind_angle==999 的数据应赋为 NA 筛掉，wind_dqc=='3'|wind_dqc=='7'的数据，wind_tcode=='9'的数据都同理过滤掉。数据处理完之后，我们要开始求解月平均风速，select 出 month 以及 wind_speed 两列数据并赋在新的 data SZ_wind_speed 中，利用 group_by 以及 summarize 计算月平均风速，然后绘制成图。



从图中我们可以看出深圳近十年月平均风速在 10 至 13 年都比较低，在 14,15 年以及 18 年左右的月平均风速比较大，总体上看，13 年开始至 20 年的风速都比较大。

Problem3

选用的水电站径流数据比较好处理，去掉 NA 值之后绘制成图

