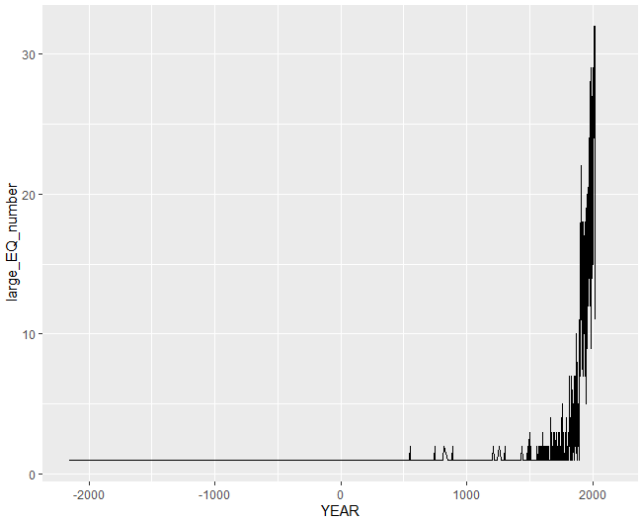
#1

```r
#1
#1.1
install.packages("tidyr")
install.packages("dplyr")
install.packages("ggplot2")
library(tidyr)
library(dplyr)
library(ggplot2)
signif <- read.csv("signif.txt", head = TRUE, sep = '\t')
Sig_Eqs <- as_tibble(signif)
#1.2
top_ten_death <- Sig_Eqs %>%
  select(COUNTRY, YEAR, DEATHS) %>%
    group_by(COUNTRY) %>%
     summarise( death_number = sum(DEATHS)) %>%
      arrange(desc(death_number))
top_ten_death[1:10,]
#1.3
Sig_Eqs %>%
  select(YEAR, EQ_PRIMARY) %>%
  group_by(YEAR) %>%
  filter(EQ_PRIMARY > 6.0) %>%
  summarise(large_EQ_number = n()) %>%
  ggplot(aes(x=YEAR, y=large_EQ_number))+
  geom_line()+
  xlim(-2150,2150)
#1.4
Sig_Eqs_NA <- Sig_Eqs %>%
  filter(EQ_PRIMARY != 'NA')
countEq_LargestEq <- function(country1){
  Sig_Eqs_NA %>%
    filter(COUNTRY == country1) %>%
    mutate(dateEQ = paste(YEAR,MONTH,DAY,sep = '-'))%>%
    select(dateEQ,EQ_PRIMARY) %>%
    summarise(EQ_Num = n(), EQ_Max = dateEQ[which(EQ_PRIMARY == max(EQ_PRIMARY))])
}
i<-1
EQlist<-matrix(ncol = 3,nrow = length(unique(Sig_Eqs_NA$COUNTRY)))
for(CountryName in unique(Sig_Eqs_NA$COUNTRY)){
  EQlist[i,]<-c(as.character(CountryName),
                as.numeric(countEq_LargestEq(CountryName)[1,1]),
                as.character(countEq_LargestEq(CountryName)[1,2]))
  i=i+1
}
#Sort in descending order by earthquake numbers.
EQlist_Order<-EQlist[order(as.numeric(EQlist[,2]),decreasing=T),]
EQlist_Order
```

1.1 用 read.csv 读取 txt 格式的文件（这样都行？！），用'\t'分格。
1.2 用国家做分组，sum 计算各个国家因地震死亡数，再用 arrange desc 降序排序，得出前十国家
1.3 以年份做分组，用 filter 筛选出大于 6 级的地震，统计地震数，用 ggplot 画图。横轴为年份（时间），纵轴为地震发生次数。
1.4 构建一个函数，去除集合中的 NA 值。在函数中通过 filter 作为条件筛选，用 mutate 加上一行年月日的信息（采用 paste），用 which 寻找该国发生的最大地震。
构造一个 n*3 的空矩阵，用 unique 求得行数用 for 循环按国家以此输入数据信息，用 order 排序（得到了李熹成同学在构造函数和 for 循环输入到矩阵的帮助）

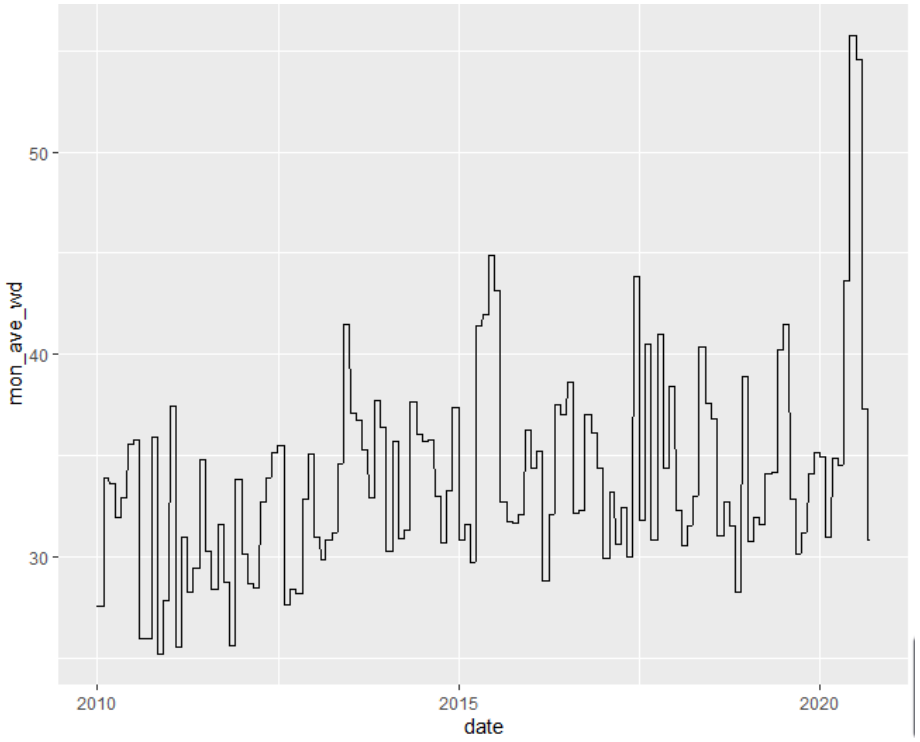
```
> top_ten_death[1:10,]
# A tibble: 10 x 2
   COUNTRY                                death_number
   <chr>                                         <int>
 1 BARBADOS                                       3000
 2 GUINEA                                          443
 3 LIBYA                                           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
```



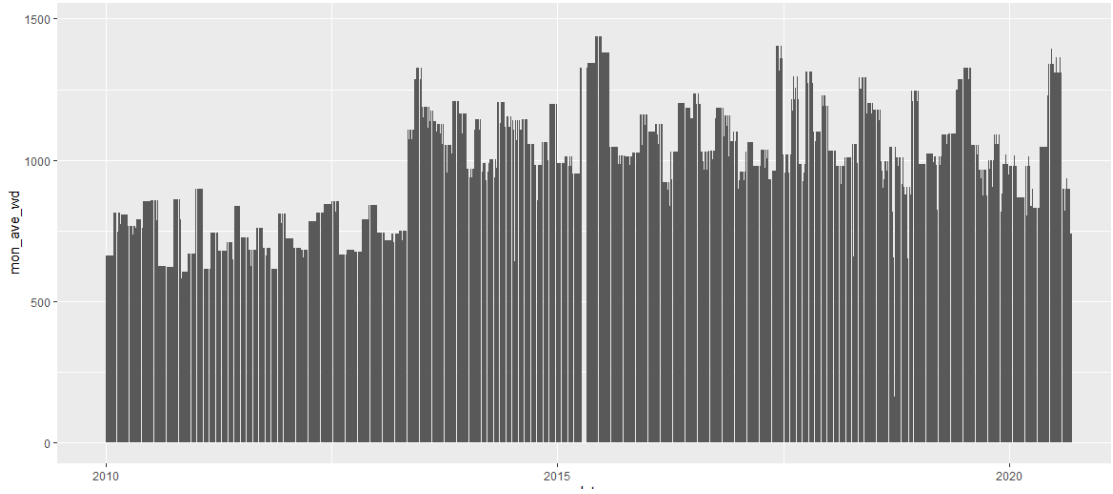

```
> EQlist_Order
      [,1]                                [,2]  [,3]
 [1,] "CHINA"                             "575" "1668-7-25"
 [2,] "JAPAN"                             "344" "2011-3-11"
 [3,] "INDONESIA"                         "314" "2004-12-26"
 [4,] "IRAN"                              "249" "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"
```

#2

```
#2
library(tidyr)
library(dplyr)
library(ggplot2)
Shenzhen_data <- read.csv("2281305.csv", head = TRUE)
Shenzhen_data %>%
  select(DATE,WND) %>%
  mutate(YM = substr(DATE,1,7), WD = as.numeric(substr(WND,9,12))) %>%
  filter(WD != 9999) %>%
  group_by(YM) %>%
  summarise(date = as.Date(DATE), mon_ave_wd = sum(WD)/n()) %>%
  ggplot(aes(x=date, y=mon_ave_wd))+
  geom_line()
```



Or use column



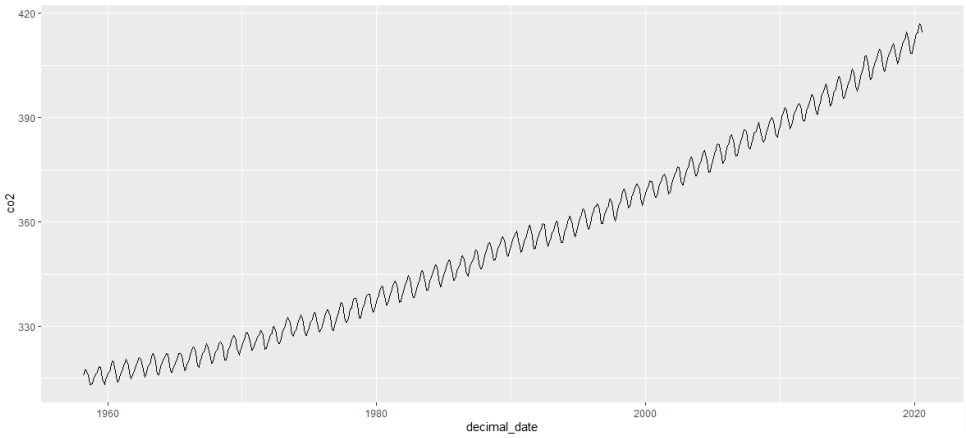
WND 为风的数据（找了好久，不是很懂 user guild 里面的指引）。选出时间和风的数据用 substr 新建年-月和风速（as.numeric 变换格式便于计算）；去除缺失的数据后按年-月（YM）分组计算月平均值，画图。 没有明显的月变化趋势，但风速变大了。

（发现没有用 as_tibble 转换，但发现结果没变，这里差别在哪里呢）

#3

（没找到合适的数据包就用了之前 CO_2 的）

```
#3
library(tidyr)
library(dplyr)
library(ggplot2)
co2_data <- read.csv("co2_mm_mlo.csv",head = T)
co2_tib <- as_tibble(co2_data)
co2_tib %>%
  filter(quality != 0) %>%
  ggplot(aes(x=decimal_date,y=co2))+
  geom_line()
co2_tib %>%
  select(month, co2, quality) %>%
  filter(quality != 0) %>%
  group_by(month) %>%
  summarise(maxco2 = max(co2), minco2 = min(co2) ) %>%
  arrange(desc(maxco2))
co2_tib %>%
  filter(quality != 0) %>%
  group_by(year) %>%
  summarise(maxco2 = max(co2), minco2 = min(co2)) %>%
  arrange(desc(maxco2))
```




```
   month     maxco2 minco2
   <chr>      <dbl>  <dbl>
 1 May         417.   318.
 2 June        416.   318.
 3 April       416.   317.
 4 March       415.   316.
 5 July        414.   316.
 6 February    414.   316.
 7 January     413.   316.
 8 December    412.   315.
 9 November    410.   313.
10 August      410.   315.
11 September   409.   313.
12 October     409.   313.
```

```
    year maxco2 minco2
   <int>  <dbl>  <dbl>
 1  2020   417.   413.
 2  2019   415.   409.
 3  2018   411.   406.
 4  2017   410.   403.
 5  2016   408.   401.
 6  2015   404.   398.
 7  2014   402.   395.
 8  2013   400.   393.
 9  2012   397.   391.
10  2011   394.   389
# ... with 53 more rows
> |
```