Problem Set 1

Baoyue Liang

2018/8/31

Problem 3

(a)

Step1: Fetch the temperature data from 2014 to 2018, unzip the zip file

```
for((i=2014; i<=2018; i++)); do
curl -O https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/${i}.csv.gz
gunzip ${i}.csv.gz
done;
```

Step2: Count the lines in each csv file, output as observations

```
for ((i=2014; i<=2018; i++)); do
echo "year ${i} has $(wc -l ${i}.csv | cut -d' ' -f2) observations"
done

## year 2014 has 34599683 observations
## year 2015 has 35233244 observations
## year 2016 has 28127149 observations
## year 2017 has 34748555 observations
## year 2018 has 20416455 observations
```

(b)

Step1: Download the ghcnd-stations.txt file and find "DEATH" in the file to get unique code of Death Valley. Step2: Find each line that contains the unique code of Death valley, 201X03(which means that the date is in March), and TMAX(which means that temprature is the maximum temprature).

```
curl -o ghcnd-stations.txt https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt
code=$(grep DEATH ghcnd-stations.txt | cut -d' ' -f1)
```

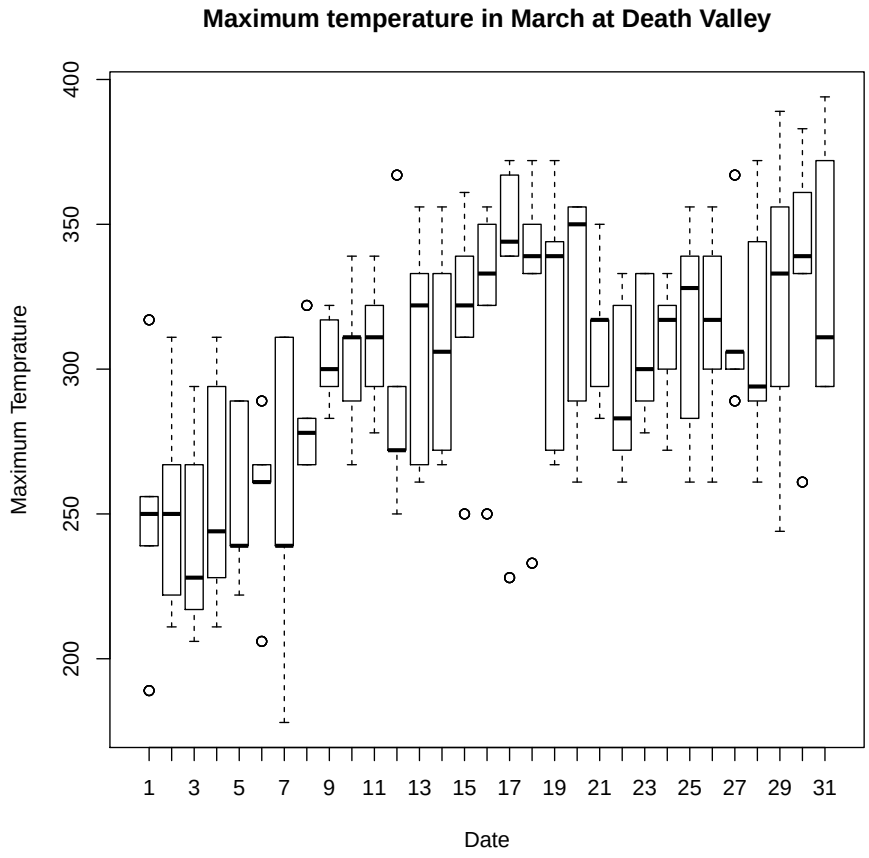
```
for((i=2014;i<=2018;i++));do
grep "${code}" ${i}.csv | grep "${i}03" | grep "TMAX" | cut -d',' -f2-4
>> TMAX.txt
done

wc -l TMAX.txt
```

```
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
  0 8959k    0 33127    0     0  14258      0  0:10:43  0:00:02  0:10:41 14254
  1 8959k    1  172k    0     0  50646      0  0:03:01  0:00:03  0:02:58 50636
  3 8959k    3  282k    0     0  68709      0  0:02:13  0:00:04  0:02:09 68708
  5 8959k    5  477k    0     0  92520      0  0:01:39  0:00:05  0:01:34 96341
  6 8959k    6  618k    0     0    99k      0  0:01:29  0:00:06  0:01:23  124k
  8 8959k    8  805k    0     0   112k      0  0:01:19  0:00:07  0:01:12  159k
 10 8959k   10  985k    0     0   118k      0  0:01:15  0:00:08  0:01:07  168k
 12 8959k   12 1149k    0     0   122k      0  0:01:13  0:00:09  0:01:04  166k
 14 8959k   14 1297k    0     0   127k      0  0:01:10  0:00:10  0:01:00  168k
 16 8959k   16 1438k    0     0   128k      0  0:01:09  0:00:11  0:00:58  164k
 17 8959k   17 1610k    0     0   132k      0  0:01:07  0:00:12  0:00:55  160k
 20 8959k   20 1797k    0     0   135k      0  0:01:06  0:00:13  0:00:53  164k
 21 8959k   21 1954k    0     0   137k      0  0:01:05  0:00:14  0:00:51  167k
 24 8959k   24 2219k    0     0   145k      0  0:01:01  0:00:15  0:00:46  182k
 28 8959k   28 2547k    0     0   157k      0  0:00:56  0:00:16  0:00:40  222k
 31 8959k   31 2782k    0     0   160k      0  0:00:55  0:00:17  0:00:38  228k
 37 8959k   37 3337k    0     0   182k      0  0:00:48  0:00:18  0:00:30  309k
 42 8959k   42 3774k    0     0   196k      0  0:00:45  0:00:19  0:00:26  367k
 46 8959k   46 4188k    0     0   207k      0  0:00:43  0:00:20  0:00:23  395k
 52 8959k   52 4696k    0     0   221k      0  0:00:40  0:00:21  0:00:19  428k
 58 8959k   58 5227k    0     0   235k      0  0:00:38  0:00:22  0:00:16  499k
 63 8959k   63 5672k    0     0   244k      0  0:00:36  0:00:23  0:00:13  474k
 68 8959k   68 6149k    0     0   254k      0  0:00:35  0:00:24  0:00:11  473k
 74 8959k   74 6665k    0     0   264k      0  0:00:33  0:00:25  0:00:08  497k
 77 8959k   77 6977k    0     0   265k      0  0:00:33  0:00:26  0:00:07  446k
 84 8959k   84 7587k    0     0   278k      0  0:00:32  0:00:27  0:00:05  471k
 90 8959k   90 8126k    0     0   287k      0  0:00:31  0:00:28  0:00:03  482k
 95 8959k   95 8579k    0     0   294k      0  0:00:30  0:00:29  0:00:01  486k
100 8959k  100 8959k    0     0   300k      0  0:00:29  0:00:29 --:--:--  498k
##      620 TMAX.txt
```

(c)

```
TMAX = read.table("TMAX.txt", header = FALSE, sep = ",")
TMAX = TMAX[,-2]
colnames(TMAX) = c("Date","Temprature")
TMAX$Day = TMAX$Date %% 100
boxplot(TMAX$Temprature ~ TMAX$Day, xlab = "Date", ylab = "Maximum Temprature")
title("Maximum temperature in March at Death Valley")
```



(d)

```
get_weather(){
  location=$1
  wea_var=$2
```

```
  month=$3
  year1=$4
  year2=$5

  if [ "$(echo $1)" == "-h" ] || [ $# != 5 ]; then
  echo "usage: get_weather location weather_type month(double digit)
year_to_start year_to_end"
  return
  fi

  if [[ "$year1" -lt "2014" || "$year2" -gt "2018" ]]; then
  echo "No data available. Easy on me."
  return
  fi

  if [ "${year1}" -gt "${year2}" ]; then
  echo -e "Please verify the year you choose. \nCurrently, we do not
offer the service to travel back in time."
  return
  fi

  mon=$(echo $month | wc -c)
  if [[ "$month" -gt "12" || "$mon" -ne "3" ]]; then
  echo -e "Please input a ligit double degit month from 01 to 12. \nThanks
for your cooperation."
  return
  fi

  curl -o ghcnd-stations.txt https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.t

  code=$(grep -i "${location}" ghcnd-stations.txt | cut -d' ' -f1)

  if [ ! "$code" ]; then
  echo -e "Unavailable weather station. \nWould you like to check your
spelling?"
  return
  fi

  if  [ "$(echo $code | wc -w)" -gt "1" ]; then
  echo -e "Please specify the location. \nJust a kind reminder, the
law states that it is illegal to marry multiple brides/bridegrooms
\neven at different places."
  return
  fi
```

```
  for ((i=$year1;i<=$year2;i++)) ;do
  #curl -O https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/${i}.csv.gz
  #gunzip ${i}.csv.gz
  grep $code ${i}.csv | grep ${i}${month} | grep -i ${wea_var} | cut
-d',' -f2-4 >> data.txt
  #rm ${i}.csv
  done
  rm ghcnd-stations.txt
  echo "Saved to data.txt, finally, please enjoy the preview"
  head data.txt
}

##Test function

get_weather "DEATH VALLEY" TMAX 06 2016 2017

##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                 Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0 8959k    0 14231    0     0   7417      0  0:20:36  0:00:01  0:20:35  7415
  0 8959k    0 61471    0     0  23151      0  0:06:36  0:00:02  0:06:34 23144
  2 8959k    2  239k    0     0  64654      0  0:02:21  0:00:03  0:02:18 64648
  5 8959k    5  458k    0     0    98k      0  0:01:30  0:00:04  0:01:26   98k
  7 8959k    7  630k    0     0   111k      0  0:01:20  0:00:05  0:01:15  129k
 10 8959k   10  903k    0     0   136k      0  0:01:05  0:00:06  0:00:59  189k
 13 8959k   13 1239k    0     0   162k      0  0:00:54  0:00:07  0:00:47  238k
 18 8959k   18 1645k    0     0   189k      0  0:00:47  0:00:08  0:00:39  288k
 22 8959k   22 2052k    0     0   214k      0  0:00:41  0:00:09  0:00:32  321k
 28 8959k   28 2575k    0     0   242k      0  0:00:36  0:00:10  0:00:26  394k
 40 8959k   40 3591k    0     0   309k      0  0:00:28  0:00:11  0:00:17  540k
 51 8959k   51 4575k    0     0   362k      0  0:00:24  0:00:12  0:00:12  664k
 71 8959k   71 6364k    0     0   466k      0  0:00:19  0:00:13  0:00:06  945k
 81 8959k   81 7302k    0     0   497k      0  0:00:17  0:00:14  0:00:03 1033k
100 8959k  100 8959k    0     0   583k      0  0:00:15  0:00:15 --:--:-- 1340k
## Saved to data.txt, finally, please enjoy the preview
## 20160601,TMAX,467
## 20160602,TMAX,489
## 20160603,TMAX,478
## 20160604,TMAX,489
## 20160605,TMAX,483
## 20160606,TMAX,461
## 20160607,TMAX,478
## 20160608,TMAX,489
```

```
## 20160609,TMAX,478
## 20160610,TMAX,444
```

Problem 4

First, we should get the url and select all the file name of the txt files. Then, we append the txt file name after the address to download the file and echo the name of the file.

```
curl 'https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/'| grep -o -E
'"\S*.txt"' > name.txt

for URL in `cat name.txt`; do
curl -O https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/${URL};
echo "${URL}";
done;
```

```
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0  6068    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100  6068  100  6068    0     0   3891      0  0:00:01  0:00:01 --:--:--  3889
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
100   249  100   249    0     0    163      0  0:00:01  0:00:01 --:--:--   163
## "ghcnd-countries.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0   249    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100   249  100   249    0     0    189      0  0:00:01  0:00:01 --:--:--   189
## "ghcnd-inventory.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
```

```
100   246  100   246    0     0    184      0  0:00:01  0:00:01 --:--:--   184
## "ghcnd-states.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0   248    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100   248  100   248    0     0    146      0  0:00:01  0:00:01 --:--:--   146
## "ghcnd-stations.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100   247  100   247    0     0    167      0  0:00:01  0:00:01 --:--:--   167
## "ghcnd-version.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0   245    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100   245  100   245    0     0    166      0  0:00:01  0:00:01 --:--:--   166
## "mingle-list.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100   240  100   240    0     0    153      0  0:00:01  0:00:01 --:--:--   153
## "readme.txt"
##   % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
##                                  Dload  Upload   Total   Spent    Left  Speed
##
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
  0   240    0     0    0     0      0      0 --:--:--  0:00:01 --:--:--     0
100   240  100   240    0     0    169      0  0:00:01  0:00:01 --:--:--   169
## "status.txt"
```

Problem 5

(b)

```
a <-c(0,1,2,3)
```

```
print r.a

## [0.0, 1.0, 2.0, 3.0]

b = r.a[::-1]
```

```
library(reticulate)
print(py$b)

## [1] 3 2 1 0
```